AD-A216 475

DOCUMENTATION PAGE

DTIC ELECTE JAN 0 5 1990 S D

1a. unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 89-1732
6a. NAME OF PERFORMING ORGANIZATION Harvard University	6b. OFFICE SYMBOL (If applicable)
7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research/NL	
6c. ADDRESS (City, State and ZIP Code) Department of Psychology 33 Kirkland Street, Cambridge, MA 02138	7b. ADDRESS (City, State and ZIP Code) Building 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) NL
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR #87-305	
8c. ADDRESS (City, State and ZIP Code) Building 410 Bolling AFB, DC 20332-6448	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102f	2313	A4	

11. TITLE (Include Security Classification)
Multi-Level Processing in Human Speech Recognition (Unclassified)

12. PERSONAL AUTHOR(S)
Peter C. Gordon

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final Technical Rpt	FROM 6/88 TO 6/89	Sept. 6, 1989	35

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB. GR.
05	09	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
speech perception, prosody, context effects, phonetic segments, fricatives.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This project has investigated the thesis that perception of the speech signal occurs at different levels of resolution. It has addressed this thesis in the domain of the temporal components of speech, where multiple levels of resolution are evident in the prosodic (macrostructure) and segmental (microstructure) levels of analysis. The body of this report is divided into three parts. The first part addresses interactions between different levels of temporal information in the speech signal. The second part addresses complexities that occur in the use of temporal cues in recognizing phonetic segments. One study in this section explores the dependencies between vowel and fricative identities that are cued by the same durational acoustic cue. A second series of studies conducted with Jennifer L. Eberhardt, explores the effects of attention on the perceptual salience of temporal cues to the identity of phonetic segments. The third part of this report, discusses work, conducted with David W. Gow, that addresses the macro-level of temporal information. This work explores the role of stress in recognition and memory

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☒ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Alfred R. Fregly	22b. TELEPHONE NUMBER (Include Area Code) (202) 767-5021
22c. OFFICE SYMBOL NL	

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

SECURITY CLASSIFICATION OF THIS PAGE
unclassified

19

90 01 04 048

Multi-Level Processing in Human Speech Perception

Final Report to the
Air Force Office of Scientific Research
Life Sciences Directorate

Peter C. Gordon
Harvard University
Cambridge, MA 02138

Project Summary

This project has investigated the thesis that perception of the speech signal occurs at different levels of resolution. It has addressed this thesis in the domain of the temporal components of speech, where multiple levels of resolution are evident in the prosodic (macrostructure) and segmental (microstructure) levels of analysis. The body of this report is divided into three parts. The first part addresses interactions between different levels of temporal information in the speech signal. The second part addresses complexities that occur in the use of temporal cues in recognizing phonetic segments. One study in this section explores the dependencies between vowel and fricative identities that are cued by the same durational acoustic cue. A second series of studies, conducted with Jennifer L. Eberhardt, explores the effects of attention on the perceptual salience of temporal cues to the identity of phonetic segments. The third part of this report, discusses work, conducted with David W. Gow, that addresses the macro-level of temporal information. This work explores the role of rythmically-based stress effects in the recognition and memorial representation of syllables.

Status of Research

Below is a list of publications and presentations supported by the grant. The remainder of the document describes additional research that has been conducted on the project.

Publications supported by the grant

Gordon, P.C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. ***Perception & Psychophysics*, 43**, 137-146.

Gordon, P.C., & Eberhardt, J.L. (1988). Effects of attention on the phonetic importance of acoustic cues. ***Journal of the Acoustical Society of America*, 84**, Suppl. 1.

Gordon, P.C. (1989). Perceptual-motor processing in speech. In T.G. Reeve & R.W. Proctor (Eds.), ***Stimulus-Response Compatibility: An Integrated Perspective.*** North Holland.

Gordon, P.C. (in press). Context effects in recognizing syllable-final /z/ and /s/ in different phrasal positions. ***Journal of the Acoustical Society of America.***

Yaniv, I., Meyer, D.E., Gordon, P.C., Huff, C.A., & Sevald, C.A. (In press). Vowel similarity and syllable structure in motor programming of speech. ***Journal of Memory and Language.***

Presentations

Induction of rate-dependent processing by coarse-grained aspects of speech. Haskins Laboratories, March 31, 1988.

Part1
Temporal Macrostructure as Context for Segment Recognition

As indicated above, an essential claim of the present view of speech perception is that listeners use the temporal macrostructure of an utterance as context for interpreting certain acoustic cues to the identity of phonetic segments, and that they can derive the temporal macrostructure from the acoustic signal without recognizing its underlying phonetic segments. We have done a variety of studies that support this claim and begin to establish the generality of this strategy of speech processing.

Many of the fine-grained acoustic cues to the identity of phonetic segments consist of the durations for which certain acoustic properties are present, or the rate at which they change. Following Nakatani et al. (1981) and Port et al. (1980), we are referring to these acoustic characteristics as the temporal microstructure of speech. Temporal cues of this sort contribute to the recognition of a large number of phonetic distinctions. In research completed to date, we have examined the perception of temporal cues to three different phonetic distinctions: closure duration as a cue to voicing in intervocalic stops, rate of first formant transition as a cue to the stop/glide distinction, and vowel duration as a cue to voicing in syllable-final fricatives.

Voicing in intervocalic stops.

In an initial series of studies (Gordon, 1988), the effect of overall speaking rate on the perception of voicing in intervocalic stops was studied. Consistent with previous studies (Miller & Grosjean, 1981; Port, 1979), it was found that the boundary between voiced and voiceless percepts along a continuum of closure durations (viz. rabid vs. rapid) shifted as a function of the speaking rate of an initial precursive phrase. The boundary shift was consistent with the idea that listeners interpreted the duration of the closure relative to the speaking rate of the precursive phrase. Shorter closure durations were needed to cue a voiceless percept when the speaking rate of the initial phrase was fast than when it was slow, indicating that temporal macrostructure, in the form of overall speaking rate, influenced the interpretation of temporal microstructure in a consistent and expected way. Furthermore, this influence was obtained even when the precursor phrase was modified so that the phonetic segments that it contained were not recognizable. The successful modifications involved severe low-pass filtering of the precursor phrase in one case (Gordon, 1988; Experiment 1), and using a sine wave modulated by the amplitude envelope of the precursor phrase in another case (Gordon, 1988; Experiment 2). The size of the effects obtained with the modified precursors did not differ from those obtained with the unmodified precursors. These results indicate that the temporal macrostructure of speech, at least with respect to overall speaking rate, can be extracted from relatively coarse aspects of the speech signal, and then used as a basis for interpreting more fine-grained temporal aspects of the signal.

Stop/Glide Distinction.

The results of a follow-up study help to establish the generality of the results described above. In this study, the influence of the speaking rate of a sentential context (the phrase "I'm trying to say") was studied on the perception of a second phonetic distinction, stop vs. glide, which also has a strong temporal component. The stop/glide distinction, as in /ba/ versus /wa/, is cued in good measure by the duration of the formant transitions. For /ba/, these transitions are usually shorter than for /wa/ (Liberman, Delattre, Gerstman, & Cooper, 1956; Miller & Liberman, 1979). Furthermore, a number of studies have shown that the interpretation of the transition duration cue is interpreted relative to the perceived speaking rate of the utterance (Miller & Liberman, 1979; Miller, 1987). A shorter transition duration is needed to cue a /w/ when the perceived speaking rate is fast than when it is slow.

The goal of this experiment, as in those described above (Gordon, 1988), was to see whether the boundary between /b/ and /w/ on a formant-transition-duration continuum would be influenced by the speaking rate of a precursive phrase. In order to help establish the generality of the phenomenon, a larger number of tokens of the precursor phrase were used. In the earlier studies (Gordon, 1988), only one pair of fast and slow precursor phrases, produced by a single speaker, had been used. For the present study, two naive speakers were asked to produce eight tokens each of fast and slow versions of the precursor phrase, for a total of 32 phrases. All of these tokens were then used in the experiment in an attempt to increase the generality of the previous finding.

The precursor phrases were analyzed and resynthesized into coarse-grained representations that preserved only the fundamental-frequency contour and the amplitude envelope of voiced energy. This was done by determining the energy of each pitch period, and then replacing that period with a sine wave of equal energy and duration. Tokens drawn from a /ba/-/wa/ continuum were then appended to the coarse-grained precursors. The /ba/-/wa/ continuum was created by varying the duration of the formant transitions from 5 msec to 50 msec in 5 msec steps using the Klatt (1980b) speech synthesizer. Two absolute syllable durations (155 msec and 225 msec) were used. This additional way of manipulating perceived speaking rate (Miller & Liberman, 1979) was designed to provide a baseline against which to assess the effect of the speaking rate of the coarse-grained precursors. A number of continua, with different F0s, were synthesized so that the tokens could be matched to the F0 of the syllable that it was replacing in the original sentence. This assured that there was no marked discontinuity between the frequency of the coarse-grained precursor and the test syllable which was appended to it.

The experiment was analyzed by determining the boundary between /b/ and /w/ for each subject in each condition (see Overview of Methods below). The speaking rate of the precursor had a significant effect on boundary location, $F(1,9) = 9.1$, $p < .02$, as did the duration of the test syllable, $F(1,9) = 41.3$, $p < .001$. The boundary shift resulting from speaking rate was 1.25 msec while the boundary shift as a result of syllable duration was 3.5 msec. The magnitude of the boundary shift resulting from different speaking rates of the coarse-grained precursor was thus approximately 36 percent of the boundary shift obtained by varying syllable duration. This relationship between the sizes of the effects due to precursive rate information and syllable duration is comparable to those that have been found with fully articulated precursors (Port & Dalby, 1982; Summerfield, 1981). Quite reasonably, it appears that listeners weigh the distant rate information conveyed by the precursor less heavily than the local rate information conveyed by the syllable duration.

These results generalize the findings of Gordon (1988) in two important ways. First, they show that effects of the speaking rate of a coarse-grained representation can be observed on a second phonetic distinction, the stop/glide distinction. Second, coarse-grained representations of a much larger sample of precursor phrases were employed, thereby reducing the chance that the previous results had depended on idiosyncrasies of the phrases employed. In addition, the experiment showed that the relationship between the size of the effect of distant rate information and local rate information was approximately the same for coarse-grained representations of speech as for normal speech (cf. Port & Dalby, 1982; Summerfield, 1981).

Syllable-final fricatives.

The next set of studies (Gordon,in press and below) again explores a new phonetic distinction, voicing in syllable-final fricatives, and shows that its temporal cues are perceived in a context-dependent fashion. This distinction, illustrated by the difference between /jus/ as in "the use" and /juz/ as in "to use", has a number of acoustic correlates (e.g., vowel duration,

vowel-offset duration and duration of frication), many of which have been found to serve as perceptual cues to the distinction. Perhaps the most important of these cues is the duration of the preceding vowel. For /s/ the preceding vowel is typically shorter than for /z/. The present studies revealed the dependence of this cue on speaking rate as conveyed by a prosodic pattern. These studies differed from the previous ones in that they did not manipulate overall speaking rate, but instead manipulated "local" speaking rate by varying the phrasal position of the test syllables. Syllables in phrase-final position are normally found to have considerably longer durations than syllables in phrase-internal position (Klatt, 1976). This phrasal-position effect is superimposed over the vowel-duration cue to syllable-final voicing, creating the potential for ambiguity in voicing identification if information about phrasal position is not available.

Two experiments (described in detail in Gordon, in press) assessed whether this potential ambiguity was real. They used a gating methodology (Pollack & Picket, 1963) in which syllables are removed from context (gated), and presented in isolation to listeners. The effect of gating on recognition accuracy reflects the importance of context in recognizing the syllable. For example, in Experiment 1 of Gordon (in press) it was found that recognition accuracy for final fricatives on syllables in context was 3.8 percent better than for gated syllables. This indicates that the presence of context does not simply bias listeners' interpretation of cues but actually improves recognition of the speaker's intended utterance.

This overall percentage, however, is not the best indicator of the importance of prosodic context in recognizing voicing in syllable-final fricatives. As discussed earlier, a principle cue to this distinction is the duration of the preceding vowel. When this cue is combined with the prosodic effect of phrase-final lengthening, two very distinct situations can occur. For example, if a voiced fricative, i.e., /z/, occurs in phrase-final position, then *prosodic-segmental congruence* exists because both segment identity and phrase position cause vowel duration to be lengthened. Similarly, /s/ in phrase-internal position involves prosodic-segmental congruence because both factors shorten vowel duration. On the other hand, *prosodic-segmental incongruence* occurs if the fricative identity and phrasal position have opposite effects on vowel duration. This happens for /z/ in phrase-internal position and for /s/ in phrase-final position. The availability of contextual information would be expected to be of greater importance when the prosodic and segmental effects on vowel duration conflict. When they do not conflict, the listener has no need to take phrase position into account, since it can only reinforce the strength of the segmental cue. When they do conflict, the listener may erroneously interpret temporal variation resulting from phrase position as relevant to segment identification. The results of the gating study support this analysis. For stimuli with prosodic-segmental congruence, the presence of context led to an improvement in recognition accuracy of 1.0 percent, while for stimuli with prosodic-segmental incongruence, an improvement of 6.7 percent was observed. The difference between these two improvement scores, 5.7 percent, can be taken as a measure of the extent to which prosodic information *per se* contributes to improvement in recognition, independent of other factors such as information about speaker characteristics, or unnaturalness of the gated stimuli.

The results discussed above were obtained using completely detailed speech as context. A follow-up to the studies reported in Gordon (in press) has examined whether a more restricted portion of the signal can also convey this information. The contexts were low-pass filtered beginning at 375 Hz and down 50 dB by 625 Hz (see Gordon, 1988 for a discussion of the information that is left over after filtering). Using these filtered phrases as contexts, the "improvement" in recognition accuracy for stimuli with prosodic-segmental congruence was -3.0 percent and for prosodic segmental incongruence was 2.8 percent. This indicates a general decrease in the baseline for the influence of context, which is presumably due to the distraction created by the filtering or to the difference in amplitude between the filtered contexts and the unfiltered target syllables. However, the difference between these context effects was 5.8

percent; $F(1,11) = 5.4$, $p < .05$. As pointed out above, this difference provides a measure of the extent to which information about phrasal position has a contextual effect on segment identity. The magnitude of this difference with filtered contexts is comparable to the difference that was found with fully detailed speech as context. This provides further support for the idea that coarse-grained aspects of the stimulus can provide important contextual information for segment identification.

Summary and Conclusions.

The above results have shown that coarse-grained aspects of the speech signal, which are themselves insufficient for recognizing phonetic segments, can provide contextual information that is important for the accurate interpretation of temporal cues to the identity of phonetic segments. The influence of coarse-grained aspects of context have been shown on three phonetic distinctions: voicing in intervocalic stops, the stop/glide distinction, and voicing in syllable-final fricatives. This effect has been shown for overall speaking rate and for local speaking rate in the form of phrase-final lengthening. The effects of coarse-grained context have been shown with boundary-shifts along phonetic continua and with gating methodology.

The above studies have focused primarily on the ability of amplitude and F0 patterns to convey contextual information about temporal macrostructure. These signal characteristics have been selected for study because they are the ones classically associated with prosody (Lehiste, 1970). At least with respect to overall speaking rate there appears to be some redundancy in these aspects of the signal. Shifts in the phonetic boundary along acoustic continua were observed with low-pass filtered speech and with just the amplitude envelope (Gordon, 1988). The low-pass filtered speech preserves the F0 contour as well as a portion of the amplitude envelope of voiced energy. The AE-only speech eliminates information about fundamental frequency and preserves the variation in total energy over time. Despite these differences, both these coarse-grained representations conveyed the same rate information.

Part 2
Processing of Temporal Cues to Segment Identity

Mutual dependencies between phonetic segments.

A major motivation for exploring the contextual role of coarse-grained information has been the view that context-independent interpretation of durational cues is not possible because the source of temporal variation, segment identity or speaking rate, can not be determined in a context-independent manner. Klatt (1980a) has referred to this as a "classic chicken-egg problem". Gordon (1988) has further argued that using the identities of neighboring segments as context for recognizing another segment is not generally reliable because the segments may show mutual dependencies. This study was designed to assess whether a chicken-egg problem truly exists when listeners are presented with the acoustic information for a only few segments. Follow-up studies will assess whether coarse-grained aspects of the extended phonetic context can help to resolve this problem.

The study examined the recognition of a pair of adjacent phonetic segments, both of which are cued in part by temporal properties of the stimulus. Specifically, it examined the recognition of a vowel, /i/, /a/, /I/, or /ae/ and the following fricative, /s/ or /z/. The distinction between the vowels of this set is cued in part by duration; /i/ and /a/ are long (tense) vowels while /I/ and /ae/ are short (lax) vowels (Ainsworth, 1972; Peterson & Lehiste, 1960). Similarly, as discussed above, the voicing distinction in syllable-final fricatives is cued in part by vowel duration; /z/ is indicated by a long vowel and /s/ is indicated by a short vowel. Based on the gating results already obtained with fricatives (Gordon, in press), it seems

reasonable to expect that confusions in consonant voicing may result from the loss of context. Research by Verbrugge and Shankweiler (1977; discussed in Miller, 1981) indicates that confusions in vowel tenseness are likely to result when identifications must be made without the temporal context in which they were spoken. The goal of this study was to determine whether there are dependencies in the perception of these segments in naturally produced speech. This study extends previous work on this topic by Mermelstein (1978) who failed to find any such dependencies in synthetic speech. By examining the perception of a large sample of natural speech, this study can assess the extent to which speakers in general produce syllables that cause dependencies in the recognition of successive segments cued by the same temporal property.

Method. The method for the study was very similar to the gating study discussed above and to Gordon (in press). Tokens of syllables varying factorially in vowel identity and final fricative identity were produced at different speaking rates. Variations in speaking rate were achieved by placement of the target syllable in phrase-final versus phrase-internal position. Ten speakers produced two instances of each of eight syllables in each of the two sentence positions. The syllables were then gated out of context and presented in a random order to 12 subjects who were asked to identify both the vowel and the subsequent fricative.

Results. Recognition accuracy for vowels is shown in Table 1 and was quite high, averaging 97 percent. There were no statistically significant sources of variation in recognition of the vowels. Recognition accuracy for fricatives is shown in Table 2 and was not so high, averaging 87.5 percent. Statistically significant sources of variation in accuracy of identifying fricatives were assessed by computing separate F ratios (F_1 and F_2) using listener and speaker as random factors. There were two significant main effects: fricative identity [$F_1(1,11) = 81.6$, $p < .001$; $F_2(1,9) = 10.6$, $p < .01$] and vowel identity [$F_1(3,33) = 21.8$, $p < .001$; $F_2(3,27) = 6.1$, $p < .005$]. Two significant two-way interactions were also obtained: phrase position and fricative identity [$F_1(1,11) = 54.1$, $p < .001$; $F_2(1,9) = 9.0$, $p < .02$], vowel identity and fricative identity [$F_1(3,33) = 17.4$, $p < .001$; $F_2(3,27) = 5.5$, $p < .01$].

Discussion. The results on vowel identification failed to confirm the previous results of Verbrugge and Shankweiler (1977) who found that confusions between long and short vowels occurred when the speaking rate information provided by context was available because the stimulus syllables had been gated. However, it should be noted that the manipulation of speaking rate in the present study differed from theirs and that the overall high rate of recognition may have produced a ceiling that obscured any patterns in the errors that were obtained.

With respect to identification of the final fricatives, the present results are consistent with those of Gordon (1989). In particular, the present results again showed a significant interaction of fricative identity and phrase position in recognition accuracy. In particular, the voiced fricative /z/ was recognized more accurately in phrase-final position than in phrase-internal position, while the opposite pattern was observed for the voiceless fricative /s/. This finding is consistent with the idea that the primary temporal cues to fricative identity are present in the vowel interval rather than the frication interval. The present results extend those of Gordon (1989) by showing that the finding generalizes across a sample of 10 speakers.

Variation in the accuracy of fricative identification also revealed a dependency between fricative identity and vowel identity as shown by the significant interaction between fricative identity and vowel identity. A planned contrast showed that recognition was most accurate when the effect of voicing on vowel duration was consistent with the inherent variation of duration due to vowel identity [$t_1(11) = 4.2$, $p < .002$; $t_2(9) = 3.8$, $p < .002$]. This effect was present for /s/ which was perceived an average of 12.3 percent more accurately when paired with short vowels than with long vowels. A corresponding effect was not obtained for the

recognition accuracy of /z/ which did not vary as a function of vowel identity. However, recognition accuracy of /z/ was very high which may have caused ceiling effects.

The present finding of a dependence of fricative identification on vowel identity contrasts with the findings of Mermelstein (1978) who studied the same issue synthetic speech. The present use of natural speech allowed us to objectively define correct answers against which to score subjects' responses. The results showed that natural speech signals may sometimes encode information about more than one segment into a single acoustic cue in such a way that listeners can not disentangle the information. Of course, the present findings were obtained with syllables that were gated from sentences that may provide a useful temporal context for interpreting cues to segment identity. In future studies, we will assess this possibility by examining the statistical dependence between vowel and consonant identifications in syllables that presented in complete sentential context and in coarse-grained representations of the surrounding context.

Effects of Attention on the Importance of Acoustic Cues

This section describes work that follow up on Gordon and Eberhardt's (1988) investigation of the effects of attention on phonetic processing. The work below was also conducted with Jennifer L. Eberhardt. Fourteen synthetic vowels varying in formant frequency and duration were presented to subjects under high and low attention conditions. Attention was manipulated by requiring that subjects perform a non-speech distractor task while simultaneously performing a speech identification task, or by requiring that subjects perform a speech identification task only. Phonetic identifications of the vowel stimuli were found to vary with the attention condition. When subjects performed the distractor task and the speech task simultaneously, duration became a more important cue to phonetic identity while the effect of formant frequency was reduced considerably. These results provide support for viewing speech perception as a controlled cognitive process.

General Introduction

On the surface, speech perception seems quite effortless and automatic. Although the information in the acoustic structure of a sound must be redefined in terms of its linguistic meaning in order to comprehend what was said, humans almost never make a conscious attempt at this transformation. It almost seems as though the acoustic structure of the sound conveys meaning instantaneously. While speech processing is usually subjectively simple, it is a very complex cognitive skill that takes place over a number of important stages (e. g. acoustic, phonetic, lexical, syntactic, and semantic); each of these stages making some important transformation on the structure of the sound. Because informational cues as to the identity of the sound are received at so many different levels, how humans encode and integrate these multiple cues becomes an interesting psychological issue.

How important one cue is relative to another cue, within a particular level, is somewhat difficult to analyze. Which cues dominate, which cues are relatively uninformative, and which cues are absolutely necessary for speech perception to take place is not always clear. Yet, how cue integration happens between stages depends somewhat on the relative importance of the cues that are analyzed within each stage. When many characteristics of the speech stimulus suggest the same speech identity, determining which characteristic was the dominant cue in the process is not a trivial task. However, if these perceptual cues give different messages, the relative importance of the cue can be determined by how the listener ultimately identifies the sound.

Many experiments have been done at the phonetic processing level of speech that simultaneously and independently manipulate two or more cues to phonetic identification. Experiments have been conducted by varying the acoustic correlates of stop consonants. Abramsom & Lisker (1985), for instance, varied the voice-onset time (VOT) and fundamental frequency (F0) of stop consonants and found that listeners primarily used VOT to determine phonetic identity. Moreover, F0 was used only when VOT cues became ambiguous. Thus, the listener's perception in the face of conflicting messages is in some way indicative of which cues serve as primary cues and which cues serve as secondary cues in speech processing.

Cue integration also depends on how information is stored. Preserving information in short-term memory stores may be difficult at times because memory store capacity is limited (Miller, 1956). Even if the information is stored, it may be largely effected by interference and decay (Massaro, 1970; Crowder, 1971, 1982; Cowan, 1984). Many researchers attempt to explain the perceptual differences in consonants and vowels in these terms. Listeners may be able to detect acoustic differences in vowels much more so than in consonants because information in the acoustic memory store is made much more assessable for vowel perception as opposed to consonant perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Fujisaki & Kawashima, 1969, 1970). This is consistent with the theory that if information held at one stage is not activated in a timely fashion, it may deteriorate (or in some other manner be made inaccessible) and ultimately effect perceptual processing.

Speech Perception and Attention

Speech perception is an implicit cognitive skill. Speech processing takes place at numerous hierarchical stages working in parallel; it is effected by short-term memory capacity, and is therefore subject to deficiencies brought on by interference and processing delays. Consequently, speech perception can be studied as a form of information processing.

Research dealing with the brain as an information processor has largely focused on factors that place limitations on the system. In addition to the capacity limitations due to memory storage, attention allocation places further restrictions on the processing system. The manner in which information is processed is somewhat dependent on the amount of attention allocated at the time of encoding. In fact, attention is required before some cognitive skills can be performed well at all (e.g. playing chess, programming a computer, writing a masters thesis). How the brain deals with these attention restrictions is yet another interesting psychological issue.

Many researchers argue that the brain deals with attentional limitations through automatic and controlled processes (Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977; Posner, 1978; Shiffrin & Dumais, 1981). Automatic processing is effortless and speedy. It can be carried out in parallel with other automatic tasks without a reduction in performance; thus it is not limited by the capacity demands of short-term memory. Automatic processing typically occurs in well developed skill behaviors where conscious effort or selective attention is not necessarily required to perform the task. Because automatic processing is relatively permanent and inflexible, once it begins, it becomes difficult to suppress or modify.

Controlled processing, however, can be extremely flexible. It can adapt quite well to unfamiliar situations. Yet in exchange for this powerful capability, processing is relatively slow and serial. It is also limited by short-term memory capacity and, therefore, conscious attention is needed for its application. Furthermore, performing more than one task that requires this type of processing may result in interference and overall performance reduction.

The automatic-control processing paradigm has been used extensively in audition research. The application of attention models to audition tends to fall under the two broad sub-classes of selective attention and divided attention. Selective attention research deals with how humans are able to attend to certain stimuli in the face of competing, irrelevant stimuli. Broadbent's filter theory has largely dominated this area of research (Broadbent, 1958). The theory states that if stimulus input exceeds the limited capacity of a central perceptual channel, only a portion of this input will be processed beyond the sensory stage. Furthermore, while unattended information is lost before reaching short-term memory, selectively attending to a portion of the input increases the likelihood that it will make it to this stage. Most subsequent theories, however, (Treisman, 1964; Deutsch & Deutsch, 1963; Neisser, 1967; Norman, 1968; Glucksberg & Cowen, 1970) have argued that, although unattended information is very fragile, it is perceived well past the sensory stage and is at least partially analyzed. Some researchers have even gone a step further by questioning whether the changes in subjects' responses as a result of selective attention, indicate a shift in perceptual processing, or merely a shift in decision criterion (Treisman & Geffen, 1967; Treisman & Riley, 1969).

When selectively listening for tones within a certain narrow band of frequencies, detecting frequency changes outside of this critical band becomes difficult (e.g. Tanner, Swets & Green, 1956). In addition, tones with frequencies outside of the critical band interfere less with the signal than do tones with frequencies within the critical band (Scharf, 1961). This research offers additional support to the theory that selective attention may effect perceptual processing.

Research on divided attention has also helped to shed light on the nature of limited capacity. Whether capacity limitations are specific to particular cognitive skills or whether they are unified across mental abilities is unclear (Navon & Gopher, 1979). Many researchers attempt to answer this question by simultaneously presenting stimuli in two modalities (e.g. the eye and the ear) and measuring subjects' responses. Treisman & Davies (1973), for example, found that for some tasks perceptual processing is better when stimuli are simultaneously presented across modalities than when presented within a single modality. Other experiments performed in this area are consistent with this hypothesis (e.g. Miller, 1981,1982), however, others are quite inconsistent (Shiffrin & Grantham, 1974; Gilliom & Sorkin, 1974; Duncan, 1980).

Attention models such as these have rarely been applied to phonetic processing yet the implications that may come from this research are far reaching (Gordon & Eberhardt, 1988). Although phonetic processing has been considered a relatively automatic task traditionally, this consideration may not be warranted. Given that audition can be affected by capacity limitations, it is not obvious why speech perception should not be. If capacity limitations effect the auditory stage, a stage which must come before phonetic processing, then it is reasonable to hypothesize that the phonetic stage may be effected as well. The phonetic stage uses information from auditory processing as input; if this information is altered due to capacity limitations, this alteration may be reflected at the phonetic processing stage.

The present study attempts to assess this hypothesis by presenting subjects with a speech task and a non-speech task to be performed simultaneously. If phonetic identification is altered while performing the speech task and the non-speech task together, then it can be concluded that speech processing on this level is not entirely automatic. Furthermore, changes in speech perception under these conditions would be consistent with the hypothesis of shared processing capacity. Moreover, if there are systematic changes in phonetic identification due to attentional processing, it will be interesting to determine whether these changes reflect corresponding changes in the relative importance of phonetic cues.

EXPERIMENT 1

The aim of this experiment was to establish what the perception of phonetic stimuli is like under optimal listening conditions. Before studying the effect of attention level, it was necessary to establish this baseline so that meaningful comparisons could be made later. The stimuli selected for this study were synthetic steady-state vowels of short (50 msec) and of long (300 msec) durations that varied from /i/ through /I/. Figure 1 shows the spectrogram for both /i/ and /I/. As is displayed in the figure, a critical factor to the phonetic identity of the stimulus is the positioning of the first three formant frequencies. If the first formant is of relatively low frequency while the second and third formants are of relatively high frequency, the stimulus will be perceived as /i/. Yet as the first formant frequency rises and the second and third formant frequencies fall, the stimulus will become increasingly characteristic of /I/. Another cue to phonetic identity is stimulus duration. When subjects are asked to identify long and short vowels with identical spectral structures, long vowels tend to be perceived as /i/ while short vowels tend to be perceived as /I/ (Stevens, et al.,1969).

The /i/-/I/ continuum at short and long durations has been the subject of much past research (Stevens, et al., 1969; Pisoni, 1973; Pisoni, 1975). Although both formant frequency and duration work as cues to phonetic identity, the formant cue seems to be primary. This formant frequency dominance has been uncovered mainly through experiments that independently manipulate formant frequency cues and duration cues. If the cues give inconsistent messages (for example, a stimulus with formant frequencies characteristic of /i/ and with a duration characteristic of /I/) the formant frequency cue will override the duration cue as to the identity of the stimulus. The relative importance of these cues under less than optimal listening conditions remains to be seen.

Past research has also shown that consonants are perceived more categorically than steady-state vowels (Liberman, Harris, Hoffman, & Griffith, 1957; Fry, Abramson, Eimas, & Liberman, 1962; Stevens, Liberman, Studdert-Kennedy, & Ohman, 1969; Pisoni, 1971). Whereas consonants can only be discriminated to the extent that they can be identified as belonging to a distinct phonetic category, vowels can be perceived in a much more continuous fashion. Many researchers attribute this to differences in the acoustic and the phonetic short-term memory stores (Liberman, et al., 1967; Studdert-Kennedy et al., 1972; Pisoni, 1973; Fujisaki & Kawashima, 1969, 1970). The auditory memory store is assumed to encode the acoustic properties of the stimulus while the phonetic memory store is assumed to encode information about the phonetic identity of the sound. Information in auditory memory is processed early and deteriorates at a rapid pace. Because this deterioration happens faster for consonants than for vowels, there is a very heavy reliance upon phonetic information to discriminate consonants. The reason for the more rapid deterioration of acoustic information may be related the transience of the cues. Consonants are characterized by rapid formant transitions whereas vowels stay relatively steady across a longer period of time, allowing more time for spectral analysis.

Consistent with this theory, it has been demonstrated that short vowels are perceived more categorically than long vowels (Fujisaki & Kawashima, 1970; Pisoni, 1971). As vowels become shorter in duration, they behave more like consonants. It will be interesting to see whether the continuous perception of long vowels holds up under less than optimal listening conditions. If less attention is paid to the stimulus at the time of encoding, auditory information may be lost at a rapid rate for both long vowels and short vowels.

METHOD

Subjects. Eight students at Harvard University served as subjects for approximately thirty minutes. Subjects were paid $3.00 for their participation. All subjects were native speakers of English with normal hearing.

Apparatus. Seven 300 msec steady state vowels and seven 50 msec steady state vowels were synthesized on the Klatt (1980b) synthesizer. For all stimuli, the first three formant frequencies were manipulated while the fourth and fifth formants remained constant at 3500 Hz and 4500 Hz respectively. The first three formants varied in seven equal logarithmic steps from /i/ to /I/. For all stimuli, the bandwidths of the first three formant frequencies were fixed at 60, 90, and 150, respectively.

The 300 msec vowels did differ from the 50 msec vowels in their rise and decay time. The rise and decay time was 50 msec for the 300 msec steady state vowels and 10 msec for the 50 msec steady state vowels. For the 300 msec vowels, the fundamental frequency fell from 125 Hz to 80 Hz while for the 50 msec vowels the fundamental frequency fell from 125 Hz to 100 Hz. These vowel stimuli are very similar to those used by Pisoni (1975) in an experiment dealing with auditory short-term memory.

Design. The experiment consisted of 16 test blocks with 21 vowel stimuli to a block. For half of the subjects the first eight blocks consisted of 300 msec vowels only while the remaining eight blocks consisted of 50 msec vowels. The remaining subjects heard the 50 msec vowels during the first half of the experiment and the 300 msec vowels during the latter half. For any given block, all seven of the vowel stimuli were randomly presented at one duration three times to give a total of 21 stimuli to a block. The inter-stimulus interval was 2 sec and the inter-block interval was approximately 5 sec. All stimuli were presented through headphones at a comfortable listening level.

Procedure. At the start of the session, the experimenter explained the task to the subjects. The subjects were told that they would hear a series of computer generated synthetic speech sounds. Subjects were told that each of the sounds would sound like /i/ as in "beet" or /I/ as in "bit". They were told that if the stimulus they heard sounded more like /i/, to circle the word "beet" on their answer sheet. If the stimulus they heard sounded more like /I/ they were told to circle the word "bit" on their answer sheet. The subjects were warned that some of the stimuli may not sound exactly like an /i/ or an /I/ to their ear but they were to judge what it sounded most like and respond accordingly. The subjects were instructed to make a response for every sound presented. After the subjects went through 8 blocks they were given a short break and told that the remaining blocks would contain sounds of a shorter (or longer) duration. Subjects were tested individually or in groups of two. The entire session lasted approximately 30 minutes.

Results

Figure 2 shows the mean proportion of /i/ responses as a function of frequency and duration. For simplicity, the stimulus number rather than the actual formant frequency values making up the stimulus has been presented. The stimulus numbers together will be referred to as the formant series continuum. As is indicated in Table 3, a low series value will represent a relatively low first formant frequency and relatively high second and third formant frequencies; thus it will be characteristic of /i/. Conversely, a high series value will represent a stimulus with a relatively high first formant frequency and relatively low second and third formant frequencies; thus this stimulus will be characteristic of /I/.

As can be seen from the graph, the proportion of /i/ responses decreased sharply as the formant series increased; $F(6, 42) = 145.27$, $p < .001$. The main effect of duration was also

significant; $F(1, 7) = 36.99$, $p < .001$. Subjects gave significantly more /i/ responses for the 300 msec vowels than for the 50 msec vowels. As is evident from the graph, although duration had a substantial effect at intermediate formant series values, this effect was drastically minimized at the end points. Thus, the interaction of formant series value and duration was significant; $F(6, 42) = 6.3$, $p < .001$.

Discussion

Although there was a significant effect of duration on the proportion of /i/ responses, this effect was minor relative to the effect of the formant series value. As is apparent in the graph, formant series seemed to be the dominant cue in determining the phonetic category of the vowel stimulus. Duration was used as a cue only when formant frequency information was ambiguous (at intermediate points on the series continuum). This finding is consistent with past research dealing with this topic.

What is inconsistent, however, is that duration seemed to have no effect on the perceived categoricalness of the vowels. As can be seen from Figure 2, relatively categorical functions were produced for both the 300 msec vowels and the 50 msec vowels. For both vowel durations, equal logarithmic changes in formant frequency produced very unequal changes in vowel perception. Moreover, these changes are directly related to the phonetic category of the stimulus.

This result may be due to the task that the subjects were asked to perform. Whereas this experiment involved an identification task, most other experiments relevant to this issue involved discrimination tasks (Liberman, et al., 1967; Fry et al., 1962; Stevens et al., 1969; Pisoni, 1973; Pisoni, 1975; Fujisaki & Kawashima, 1969, 1970). Discrimination tasks require subjects to detect small differences in sounds, while identification tasks merely require subjects to group these changes under a phonetic label. In this way, discrimination tasks are a more sensitive measure of perception than identification tasks. However, a discrimination task was not chosen for the present experiment because of the attention variable to be added in Experiment 2. Asking the subject to discriminate between two or three different sounds and to simultaneously perform a distractor task seemed a bit too complicated. Yet, because the method chosen does not make differences in categoricalness apparent, this issue will be dropped from further discussion.

EXPERIMENT 2

The goal of this experiment was to assess the perceptual changes of phonetic stimuli due to attentional processing demands. Attention was manipulated in this experiment by requiring performance on an arithmetic distractor task in addition to identification of a phonetic stimulus or requiring a response to the speech stimulus only. Gordon & Eberhardt (1988) have performed a similar experiment by assessing the effects of attention on stop consonants. A non-speech task performed simultaneously with a speech identification task, had a large effect on consonant identification. Moreover, by manipulating VOT and FO, they found that while VOT was the dominant cue to phonetic identity under optimal listening conditions, it became less important while attention was divided. Conversely, FO became more important in a divided attention condition as compared to an optimal listening condition. The results of the present experiment will be used to further examine the general processing capacity hypothesis and to provide further demonstration of the relative importance of cues.

METHOD

Subjects. Twelve students at Harvard University served as subjects for approximately one hour. Subjects were paid at a base rate of $4.00 but could earn additional money (up to $3.00) depending on how well they performed the arithmetic distractor task. All subjects were native speakers of English with normal hearing.

Apparatus. The vowel stimuli from Experiment 1 were employed. An IBM personal computer was used to present the distractor stimuli to the subject, to provide feedback on the distractor task, to present speech identification options to the subject, as well as to record subject responses.

Design. This experiment consisted of 8 experimental test blocks with 35 vowel stimuli in a block. Each block consisted of seven steady state vowels of one duration (300 msec or 50 msec). In a given block, all seven stimuli of a particular duration were presented five times to give a total of 35 trials to a block. In 4 of the experimental test blocks the vowel stimuli were presented while the subject was performing an arithmetic task; this is called the *distractor condition*. Stimuli in the remaining 4 experimental test blocks were presented under optimal listening conditions; this is called the *no-distractor condition*. For half of the subjects, the first two experimental blocks consisted of 300 msec vowels. The first block was presented in the distractor condition and the second block was presented in the no-distractor condition. In the third and fourth blocks 50 msec vowels were presented in the distractor condition and the no-distractor condition, respectively. The remaining blocks were identical to the first 4 blocks. Stimuli were presented to the other half of the subjects in the same fashion, however, they heard the 50 msec vowels first. All stimuli were presented through headphones at a comfortable listening level.

Procedure. At the start of the session, the experimenter read the instructions to the subjects. The subjects proceeded to do a practice block of 35 trials. The practice block involved performing the distractor task without the accompanying vowel stimuli. The subjects were told to click a mouse button to begin each trial. At the start of each trial two lines appeared on the computer screen to warn the subjects of an upcoming visual stimulus. These fixation lines were shown for approximately one second and then the visual stimulus appeared. For the practice block and for all experimental blocks in the distractor condition, the visual stimulus consisted of three numbers which were all multiples of ten. The subjects were asked to decide whether these numbers had equal differences between them or not as quickly and as accurately as possible. The subjects were told to respond by clicking the appropriate mouse button on the computer. The number of trials requiring affirmative and negative responses were roughly equal for each block. The subjects were given feedback on accuracy, speed, and points earned immediately following the subjects' response to the number task. Points were based on both the accuracy and the speed of the response. For accurate responses, as the subjects' reaction time decreased, the number of points earned increased. For inaccurate responses, negative points were earned regardless of reaction time. Subjects were paid bonus money at the end of the session according to how many total points they obtained.

After the practice block, subjects began their first experimental test block in the distractor condition. The subjects were told that their task would be the same as it was in the practice block but this time, as they performed the number task, they would hear some computer generated synthetic speech sounds. One vowel stimulus was presented per trial. The subjects were told that the auditory stimulus would sound like /i/ as in "beet" or /I/ as in "bit". After the subjects made a response to the number task, they were prompted to respond to the sound they heard during the number task by the appearance of an "e" and an "i" on the computer screen. At this time the subjects were asked to decide whether the sound they heard sounded

more like an /i/ or an /I/ by clicking the appropriate mouse button. The subjects were told to respond as accurately as possible to the auditory task although speed was not important. The subjects were told that bonus money would still be allocated according to how well the arithmetic task was performed. Therefore, the subjects were instructed to treat the arithmetic task as primary and the auditory task as secondary. Feedback was given for the arithmetic task only.

The next experimental test block was presented in the no-distractor condition. In this condition, three pairs of zeros appeared on the computer screen as the vowel sound was presented. The length of time that the visual stimulus was displayed was derived from the subjects' average response time to the number task in the previous distractor condition. After the visual stimulus left the screen, the subjects were prompted to respond to the auditory stimulus by an "e" and an "i" appearing on the computer screen. Subjects were again told to try to respond accurately, although speedy responses were not necessary. Figure 3 shows the sequence of events for a given trial.

The blocks alternated from the distractor condition to the no-distractor condition for the remainder of the session. At the end of each block, the experimenter interacted with the subjects to warn them of the nature of the upcoming block, to check their performance on the distractor task (when appropriate) and to encourage them to do better. After 4 experimental blocks elapsed, the subjects were given a short break. Each subject was tested individually for approximately one hour.

Results

Figure 4 shows the mean proportion of /i/ responses as a function of attention condition, vowel duration, and formant series. For, the reader's convenience, these results are also reported in Table 4. A within-subjects analysis of variance revealed that the main effect of attention condition was significant [$F(1, 11) = 5.04$, $p = .04$] as were the main effects of vowel duration and formant series [$F(1, 11) = 47.4$, $p < .001$ and $F(1, 11) = 235.9$, $p < .001$, respectively].

As can be seen from the graph of these results in Figure 4, the proportion of /i/ responses as a function of vowel duration and formant series in the no-distractor condition looks very similar to the results obtained in Experiment 1. This demonstrates that the results obtained in the no-distractor condition are equivalent to the results that would have been obtained under traditional listening conditions.

In the distractor condition, however, there was a significant shift in the proportion of /i/ responses as a function of formant series. Figure 5 shows that the probability of responding /i/ when presented with a sound at the low end of the series continuum is smaller in the distractor condition than in the no-distractor condition. Conversely, the probability of responding /i/ when presented with a sound at the high end of the series continuum in the distractor condition is much larger than in the no-distractor condition. Consequently, the interaction of attention condition by formant series reached significance; $F(6, 66) = 17.33$, $p < .001$. This result demonstrates the minimized importance of formant frequency as a cue to phonetic identification due to the distractor task.

Figure 4 also makes apparent the significant effect of duration. In both attention conditions, the proportion of /i/ responses was greater for the 300 msec vowels than for the 50 msec vowels. What is even more interesting is that duration had an even greater effect in the distractor condition than it did in the no distractor condition; $F(1,11) = 13.42$, $p < .004$. Figure 6 makes this interaction clear. This figure shows that the difference in the proportion /i/

responses as a function of duration was greater in the distractor condition than in the no-distractor condition.

There was also a significant interaction of duration by series; $F(6, 66) = 18.16$, $p < .001$. This interaction, however, is somewhat misleading. Because the 300 msec vowels showed a much greater effect of attention condition at the higher formant series values than at the lower formant series values, while the attention condition effected the 50 msec vowels in a much more proportional manner, the magnitude of response differences by duration is larger at higher formant series values than at lower formant series values. Because this result is largely due to a third variable (attention condition), it would much more informative to interpret the result in relation to the three-way interaction.

The interaction of attention condition, vowel duration, and formant series was significant; $F(6, 66) = 2.57$, $p = .027$. Figure 7 shows the difference in proportion /i/ responses for the 300 msec vowels and the 50 msec vowels. This difference is shown as a function of attention condition and formant series. It is interesting to note that whereas duration has a very minimal effect at very low or at very high formant series values in the no-distractor condition, the effect is much larger in the distractor condition. As a result the planned contrast was significant; $t(11) = 3.57$, $p < .005$. It seems as though duration began to play a role in phonetic identification even when cues to formant frequency were unambiguous. This is consistent the findings from the Gordon & Eberhardt (1988) study.

The effect of the distractor condition across experimental test blocks is also quite interesting. Although the proportion of accurate responses to the distractor task remained above .93 across all blocks [there were no significant changes in the proportion of accurate responses as a function of experimental block; $F < 1$], the average reaction time to the distractor task did change significantly as the session progressed; $F(1, 11) =$, $p < .018$. Figure 8 shows that the mean reaction time to the distractor task across all subjects was 1535 msec in the first experimental test block but dropped to 1363 msec by the last experimental test block. Because this speed increase was not accompanied by a decrease in accuracy, it is obvious that the distractor task became easier to perform as the session progressed.

Given this, it is reasonable to hypothesize that as the distractor task difficulty level decreased, the amount of processing required to perform the task decreased as well. Furthermore, because the amount of processing required to perform the task was reduced, the amount of attention allocated to the distractor task was reduced also. If this hypothesis is true, responses to the speech stimuli in latter experimental test blocks should more closely approximate responses obtained under optimal listening conditions.

Figure 9 shows the proportion of /i/ responses as a function of attention condition, vowel duration, and formant series. The top panel shows data from the first half of the session while the bottom panel shows data from the second half of the session. As can be seen from the figure, the attention effect was minimized considerably in the second half of the session. To determine whether the effects of session half were due to a change in the attentional demands of the distractor task, the three-way interactions of attention by duration by half and of attention by formant series by half were computed. Both interactions were determined to be significant [$F(1, 11) = 5.6$, $p < .034$ and $F(6,66) = 4.23$, $p < .001$, respectively]. Although duration cues became more important in the distractor condition across the entire session, duration had an increased impact during the first half of the session. Likewise, although the reduced importance of formant series is apparent in the entire session, this reduction is much more apparent in the first half of the session.

The effect of session half, may be due to the subjects' increased performance on the distractor task as a function of practice. The significant four-way interaction (attention x

duration x formant series x half) certainly offers support for this hypothesis [$F(6, 66) = 2.7$, $p < .021$]. Whereas duration is having a larger impact at extreme formant series values in the distractor condition as compared to the no-distractor condition, this effect is dependent in part on the session half.

Discussion

Let us return to Figure 4 which represents the mean proportion of /i/ responses as a function of attention condition, vowel duration, and formant series for the entire session. Although the figure makes the effect of attention condition apparent, one could still argue that these effects are due only to additional random responding by the subjects to the vowel stimuli in the distractor condition. After all, if the subjects' primary task is to attend to arithmetic calculations, and feedback is given on this task only, then the subjects' performance on the speech task may have simply decreased overall. This general performance decrement may be reflected in the greater proportion of speech identification responses that cluster around the 50% chance level.

However, upon closer examination it becomes apparent that the performance on the speech task in the distractor condition moves closer to chance only when considering the formant series factor alone. Moreover, when the duration factor is examined separately, performance on the speech task moves even further away from chance in the distractor condition. There is a heavier reliance on this factor as a cue to the phonetic identity of the speech stimulus. In addition, although formant series diminishes in importance for both vowel durations in the distractor condition, it diminishes even more so for the long duration vowels. This differential effect is made much more apparent in the first half data shown in Figure 9. The effects of attention on the identification of speech stimuli is quite robust, however practice on the distractor task reduced these effects somewhat. Experiments that involve a distractor task that does not change in its attentional demands need to be carried out in the future.

Experiment 2 was performed in an attempt to answer two major questions; 1) Can the processing of speech be affected by the simultaneous performance of another non-speech task? and 2) Does attention have an effect on the relative importance of characteristic cues to speech identity? The results of this experiment suggest that speech processing can be effected by the performance of another non-speech task. Human processing capacity then, seems to be at least somewhat unified across cognitive skills. However, although the non-speech task chosen for this experiment did not involve the processing of speech, it may have still involved the accessing of a verbal memory store. The interference of the distractor task with speech processing could be due to interference taking place at this level. Further experiments involving a more clear cut, non-verbal distractor may be needed.

The phonetic importance of cues also changed as a function of attention to the speech stimulus. However, why vowel duration became more important in the distractor condition as formant series became less important (and not the reverse for example) is unclear. One hypothesis is that duration is more of a salient cue than is formant frequency. If it is a salient cue, it may be relatively easy to encode. Conversely, greater effort may be required to perform the more detailed formant frequency analysis. Because in this experiment, a speech task and a non-speech attentional demanding task were performed simultaneously, less processing capacity could be devoted to speech analysis. Because processing formant frequencies for vowels takes more attentional requirements than the processing of vowel duration, a heavier reliance was placed on vowel duration as a cue to speech identity. If the processing of vowel duration is more automatic, it will be less effected by capacity limitations, and it may take over when more controlled mechanisms such as formant frequency analysis, can not be accessed.

Why the shift in the phonetic importance of these cues happened as it did, cannot be determined from the results of the present experiment. However, developmental research as well as research on hearing impaired listeners may help to resolve this issue. The phonetic importance of cues has been shown to vary as a function of age (e.g. Bernstein, 1983; Price & Simon, 1984) as well as hearing ability (e.g., Lindholm, Dorman, Taylor, & Hannley,1988). To begin to understand which cues are automatic, and thus which cues are not effected by attentional processing, it may be necessary to understand how these cues develop. In other words, developmental research may help to shed light on which speech processing skills are built-in, automatic mechanisms and which skills are learned and may, therefore, be altered through practice or attention. Research with hearing impaired listeners may also be valuable because changes in speech identification brought about through the inability to process certain perceptual cues, helps to determine which cues are vital or dominant in speech processing.

GENERAL DISCUSSION

The results of this study suggest that speech processing may not be as automatic or effortless as it seems. Conscious attention to some extent determines how speech sounds will be perceived. The fact that high level cognitive skills may have a direct impact on the low level processing of phonetic cues provides support for viewing speech perception as a controlled cognitive process.

The present experiment has examined the effects of attention at the phonetic stage, however, attention may effect processing at more than one stage (e.g. auditory, lexical, or semantic). Transformations made on the speech stimulus within each stage as well as perceptual information shared between the stages, may be effected by attention. Research on the degree to which these other stages are dependent on attentional processing would begin to make apparent the magnitude of attention effects in speech processing.

Whether the results in the present experiment are due to a change in perceptual encoding or to a change in the integration of information processed at the auditory and phonetic stages, is not immediately apparent. One way to begin to answer this question would be to fit the data to a mathematical model (Oden & Massaro, 1978) . Modeling may help to determine how cues are registered and combined by quantitatively mapping perceptual cues to phonetic importance. With a model, cue significance can be examined under varying levels of attention and the resulting differences can be assessed quantitatively. More data is needed to make model application useful to the present experiment, however, this approach may prove fruitful for future experiments.

There is one final point worth making. More often than not, experiments that investigate speech perception do so under optimal listening conditions. However, outside of the laboratory, speech processing does not always occur under such conditions. In fact, the processing of speech in more natural settings may occur at a variety of attention levels. As has been indicated by the present study, what is an important characteristic cue to speech identity at one attention level, may be drastically minimized at another. This finding may considerably reduce the generalizability of many speech processing theories.

Part 3
Prosodic Effects on Syllable Recognition

This section of the report discusses work, conducted with David W. Gow, that addressed the macro-level of temporal information. Its primary goal was to determine whether the effects of phonological stress on syllable monitoring and a retrospective probe task differ. Two experiments were performed to this end. The first was a syllable monitoring task in which the predictability of the syntactic category (and thus, the placement of primary word stress) of a target-bearing word was manipulated. This revealed significant reaction time facilitation for the detection of stressed syllables and syllables appearing in words with highly predictable syntactic categories. A second experiment employing the same stimuli and manipulations was also performed. This experiment used a retrospective probe task which required subjects to determine whether or not they had heard a target syllable in a sentence presented immediately prior to the probe. This paradigm yielded a significant reaction time facilitation effect for stressed syllables, but no effect for syntactic predictability. The results of the two experiments are taken together as evidence that anticipatory and retrospective stress effects are mediated by different mechanisms. The implications of the retrospective probe paradigm for examining the psychological reality of Liberman and Prince's (1977) theory of metrical phonology are discussed.

1. General Introduction

Phonological stress is notoriously difficult to measure. While syllable duration, fundamental frequency, intensity and formant structure may frequently vary with stress, there is no general consensus on the relationship between acoustic variables and phonological stress. Fry (1958) argued that stress is the result of complex, context-dependent interactions between these acoustic variables. Lieberman (1965, 1967) on the other hand, claimed that stress is assigned by the listener on the basis of acoustic cues, and the listener's implicit knowledge of phonological rules. Similarly, Chomsky and Halle (1968) noted that while there are physical correlates of stress, these correlates by themselves cannot account for the full range of stress distinctions experienced by the native listener. This lack of consensus leaves researchers without a clear empirical definition of stress. This means that psychologists studying stress-related phenomena must begin with the rationalist criteria of the linguist in constructing and analyzing stress features in speech samples.

Despite the lack of an physical understanding of what stress is, or how it can be measured, there is a growing literature on role of phonological stress in speech recognition processes. In recent years Cutler and Norris (1988), and Grosjean and Gee (1987) have proposed that a regular alternation of stressed and unstressed syllables in continuous speech is used to guide word segmentation and lexical access in speech perception. This is a particularly attractive notion because it suggests a viable mechanism for feedforward or anticipatory processing - a central problem in time-limited complex cognitive processes (Hebb, 1949).

The idea that rhythmic alternations in stress structure speech, and in turn may be used to guide speech perception is not a new one. Sweet (1908) made this argument about phonological stress, and extended it to non-linguistic domains including music perception. Sixty years later, Chomsky and Halle (1968) noted regular patterns in the alternation of syllabic stress in English. Martin (1972) proposed that stress alternation is used to predict content words in speech. This led to work by Shields, McHugh, and Martin (1974), and Cutler (1976) which demonstrated that pretarget prosody facilitates the processing of stressed syllables in phoneme monitoring tasks.

Current interest in stress alternation can be attributed in large part to the phoneme monitoring literature, as well as work done in the late seventies in the emerging field of metrical phonology (Liberman, 1975;Liberman and Prince, 1977; Selkirk, 1984). Metrical phonology provides sophisticated, well- articulated linguistic models for the assignment of syllabic stress and the preservation of regular stress alternation. The appropriation of ideas from metrical phonology for psycholinguistic theory-building is part of a broader movement toward the integration of linguistic and psychological theories. Recent work as Pinker's (1984) examination of linguistic constraints in language acquisition and Grodzinsky's (1986) syntactic analysis of agrammatism provide provocative examples of how linguistic theories can guide and inform empirical research in psychology. The formal rigor of linguistic theories promises to provide psycholinguistic models with a new level of sophistication if this melding of insights can be attained.

We would like to argue though, that at least in the domain of stress-related phenomena, the sophistication and formalism of current theories in both psycholinguistics and linguistics contrasts sharply with our inability to formally determine what stress is, or how it should be measured. Cooper and Eady (1986) for example, have shown that there is room for disagreement about the form of the basic data used by linguists. They had phoneticians who were naive to Liberman and Prince's (1977) theory make syllable stress judgments based on sentences that were used as central evidence in Liberman and Prince's paper. They found that the ratings of their subjects did not conform to those of Liberman and Prince, or to the predictions made by Liberman and Prince's theory. Unfortunately, this conflict over subjective ratings is not resolved by physical analysis of actual speech. Looking at the effects of Liberman and Prince's syntactic manipulations on syllable timing, Cooper and Eady (1986), and Rackerd and Fowler (1984) arrived at correspondingly conflicting interpretations. These results underline the difficulties inherent in examining stress phenomena without a clear understanding of what stress is, and how it can be examined empirically.

Given the lack of a simple or consistent physical correlate of stress, one must find a task which is sensitive to peoples' perception of stress, if one is to study it. In the realm of continuous speech perception, phoneme monitoring provides the best currently available paradigm. Several studies have used phoneme monitoring tasks to demonstrate the value of stress alternation in anticipatory processing. Shields, McHugh and Martin (1974) had subjects perform a phoneme monitoring task using target-phoneme-bearing nonsense words inserted into noun positions in spoken sentences. They found that subjects responded faster to targets placed at the beginning of the first syllable of the nonsense word if that syllable was stressed than they would if it was unstressed. When these words were edited into the context of a string of nonsense words this effect disappeared. This lead Shields et al. to conclude that rhythmic cues in the pretarget context were responsible for the stress effect. This interpretation was supported by the results of several studies. Buxton (1983) found that serial position effects associated with stressed syllable response facilitation in a phoneme monitoring task occurred given both normal and prosodically intact but semantically nonsensical speech ("jabberwocky"). This suggests that facilitation effects are not dependent on subjects using semantic information to anticipate targets. Cutler (1976) directly examined the role of rhythmic context in this effect. She manipulated pretarget rhythmic context by cross-splicing, and found that subjects responded faster to targets in unstressed syllables given a context taken from a sentence with a stressed syllable in the same position, than they did to stressed syllables following non-stress predicting contexts. Following the suggestion of Martin (1972), Cutler hypothesized that this reflects a predictive role of stress. She suggests that the rhythmic qualities of stress are used to anticipate stress bearing syllables. This prediction allows one to focus processing resources on specific segments of the acoustic stream before they are actually perceived.

Cutler and Foss (1977) offered a suggestion for why it should be useful to be able to anticipate stress. Subjects performing phoneme monitoring tasks tend to respond faster to

word initial phonemes in content words (such as nouns or verbs) than they do to the same targets in function words (such as prepositions or conjunctions). Chomsky and Halle (1968) pointed out that, unlike content words, function words do not take stress. Cutler and Foss argue that the predictable alternations of stress can be used to focus attention on content words, which they believe bear more information than function words. To demonstrate the role of stress alternation in this content/function word phoneme monitoring effect, they performed a cross-splicing experiment in which predictive stress information was shown to be responsible for the content word advantage.

In assessing this literature, it is important to distinguish between stress-related, and non-stress-related response facilitation effects. The phoneme monitoring task was originally devised to examine syntactic complexity (Foss and Lynch, 1969). Subsequent work has shown that detection in monitoring tasks is facilitated when target words are highly predictable based in preceding context (Morton and Long, 1976), or they are immediately preceded by high frequency (Foss, 1969) words. Phoneme detection latency has similarly been shown to be increased when target phonemes are preceded by ambiguous words (Foss, 1970), or non-words (Cutler and Norris, 1979). Foss (1969) argued that phoneme monitoring effects reflect the processing load placed on a limited-capacity central processor serving sentence comprehension. More recently, Cutler and Norris (1979) have surveyed the monitoring task literature, and suggested that these different effects are attributable to different components of the comprehension process. In either case, it is clear that one must be able to differentiate between stress and non-stress monitoring effects, if one is to use the monitoring task to operationalize stress.

EXPERIMENT 1

The first experiment is an attempt to disentangle stress effects from those of other forms of contextual information in a monitoring task. To this end, syntactic predictability, target position within a word, and syllable stress were concurrently manipulated in a syllable monitoring task.

Method

Materials. Each subject heard 120 sentences in the course of testing. All sentences were recorded from the reading of a male speaker of Standard American English, and then computer digitized at a sampling rate of 10 KHz. The reader read each sentence to himself before reading aloud to ensure normal intonation across the entire sentence.

Of the 120 sentences, 96 were distractors, and 24 were experimentally constructed to examine category-dependent stress pattern contrasts. There is a class of latinate words which appear as either verbs or nouns, which have different stress patterns depending on their category. These include words such as "conflict", which take stress on the first syllable when they appear as nouns, and take stress on the second syllable when they appear as verbs. Twenty-four such two syllable words showing a clear stress contrast were used. Each word appeared in the following in four sentence conditions: ambiguous verb, ambiguous noun, unambiguous verb, and unambiguous noun (table 5). It should be noted that the contrast between nouns and verbs is construed as a contrast in which syllable receives stress, rather than a syntactic contrast. In the ambiguous conditions, sentences were constructed in which the category of the target word could not be discerned based on the context of the words preceding it. The same pre-target-bearing word context was used for both noun and verb versions of each target word. In the unambiguous conditions, the context preceding the target word constrained its category to a single option (noun in one condition, and verb in the other condition).

The 120 trials each subject heard were divided into 6 blocks of twenty trials each. The first three blocks consisted entirely of distractor sentences to allow the subjects to become familiar with the task and establish stable response rates. The last three blocks each included 8 experimental trials. These eight included four trials probing for first syllable targets in each of the four sentence conditions, and four trials probing for the second syllable in each of the same conditions. No subject saw the same target-bearing word in more than one condition.

Subjects. The subjects were 32 graduate and undergraduate students recruited from the Harvard University community. Half of the subjects were male and half were female. All were native speakers of Standard American English with no discernable auditory or (uncorrected) visual deficits. Subjects were paid $5.00 for their participation in the study.

Procedure. Subjects were tested in a sound attenuating testing chamber, while seated at a desk. Subjects wore stereo headphones, and were seated roughly 30 inches from a CRT screen placed at eye level. Prior to each trial, they were prompted via the CRT screen to press the left button on a mouse placed on their desk. Pressing this button caused the computer to present the probe syllable on the screen. This probe was spelled out using its conventional orthography in lower case letters. In 90% of the trials (including all of the experimental trials), the probe was a two to seven letter syllable which appeared in the sentence. A small number of negative trials were also included to minimize disproportional vigilance at the ends of test sentences. This probe remained on the screen until the subject either signalled a response on the monitoring task or initiated the next trial.

Once the probe appeared on the screen the subjects were to read it to themselves, thinking of it in terms of how it would sound, rather than in terms of its orthography. Four seconds after the presentation of the visual probe, the subject heard the digitized audio stimulus over the headphones. They were to listen to the sentence carefully, and press the left button on the mouse as soon as they heard the target syllable. To ensure rapid response and minimize variance, subjects were instructed to have their hand on the mouse, and have a finger poised over the response key at all times. In the case of negative trials, they were not to make any response until they saw the prompt to initiate the next trial. The presentation of the audio stimulus was halted when the subject made her response. Reaction time data was then assembled by subtracting the latency between the beginning of the audio stimulus and the onset of the target syllable (as determined by auditory and visual inspection using a waveform editor), from the latency between the beginning of the sentence and the subject's response. Reaction times shorter than 500 msec. were discarded due to concern that they were the result of anticipations, and not the detection of target phonemes. Similarly, reaction times greater than 2000 msec were discarded due to concern that they were the result of subjects reprocessing the entire sentence after its presentation had ended. This placing of limits on acceptable reaction times is consistent with other work in phoneme monitoring (Cutler, 1976).

Results

Subjects performed the syllable monitoring task with acceptable accuracy. Monitored syllables were successfully detected on 88.66% of the experimental trials. Despite this general accuracy, failure to detect syllables, and the exclusion of outliers were responsible for the creation of two cells in which individual subjects did not contribute to specific conditions. One subject failed to respond accurately to all three sentences in which they were presented with stressed (first) syllable targets in ambiguous nouns. Another subject failed to respond accurately on three trials involving unstressed (second) syllable targets in unambiguous nouns. These cells were filled in with cell means based on the responses of the other three subjects who were given these particular stimuli.

Mean reaction times were computed for each subject for each condition. These are presented in Table 6. Two three-way ANOVAs were performed on the data. One ANOVA was run by subject, and another was run by target-bearing word. There was no main effect for stress in either comparison, $F(1,31) = 0.794$, $p = .320$ by subject, and $F(1,23) = 0.094$, $p > .500$. The ANOVA performed by subject did reveal two main effects. There were significant main effects for the position of stressed syllables in the target bearing word, $F (1,31) = 4.253$, $p = 0.045$, with targets in nouns (with the first syllable position stressed) being detected faster than targets in verbs (with the second syllable position stressed). This effect was also significant when the analysis was performed by word, $F(1,23) = 27.326$, $p < 0.001$. There was also a significant main effect for ambiguity, $F(1,31) = 1\ 4.132$, $p = .001$, in which reaction times were faster when subjects knew whether to expect a noun or a verb, than they were when the category of the target bearing syllable was underdetermined by prior syntactic information. This effect was not significant in the analysis by word, $F(1,23) = 2.832$, $p = 0.102$.

In addition to these main effects there was one significant interaction in the analysis by subject. The syntactic category of the target bearing word interacted with stress, $F(1,31) = 72.507$, $p < 0.001$. In verbs, stressed syllables were detected faster than unstressed syllables, while in nouns, unstressed syllables were detected faster than stressed ones. This means that the stress effect for syllable detection only held when the stressed syllable was word initial. There was no stress effect for second syllable targets. This interaction was not significant in the analysis by word, $F(1,23) = 0.008$, $p > 0.500$.

Discussion

The results of this experiment suggest that phonological stress does not completely account for syllable monitoring effects. This is consistent with Cutler and Norris's (1979) assertion that monitoring facilitation effects may reflect the effects of different types of contextual information on the functioning of several different sentence comprehension processes. The central question raised by the current results is why this procedure failed to produce the stress effect found by previous researchers (Shields et al.,1974; Cutler, 1976). Previous experiments attempting to distinguish between the effects of stress alternation and other types of stimulus information which could lead a subject to anticipate a probe have differed from the current experiment in two respects. They have all probed for word initial-targets, and they have all depended on manipulations such as cross-splicing or temporal displacement which create a discontinuity between the target and its preceding sentential context. Using unaltered stimuli, and probing for both first, and second syllable targets, this experiment has demonstrated that syllable detection latency is critically effected by variables other than contextual stress alternation.

The lack of a main effect for stress is explainable by the manipulation of target position used in the present experiment. While previous experimenters have all probed for word initial targets, we have probed for both word-initial, and word-final targets. An examination of the stress by position interaction means, shows that there was a stress facilitation effect for first syllable targets. This finding is consistent with the work reviewed in Cutler and Norris (1979) on stress effects in phoneme monitoring tasks. The introduction of second syllable targets balances this effect by showing faster reaction times for unstressed word-final syllables. The advantage for unstressed second syllables is a function of the stimuli used. All words with unstressed second syllables have stressed first syllables. It appears that subjects are primed for the second syllable by preliminary lexical access based on the first syllable. This idea is consistent with Cutler and Norris' (1988) assertion that strong syllables trigger segmentation of the speech signal, and prompt an initial attempt at lexical access using the strong syllable to define the beginning of a potential lexical item. This attempt at lexical access might also be strengthened by the presence of the probe in short term memory. The subject has available all of the phonological elements to needed to attempt lexical access prior to the presentation of the

second syllable. Thus, lexical information could be used in a top-down function to facilitate the detection of an unstressed second syllable target.

Cutler and Norris' segmentation hypothesis also explains the reaction time advantage that was obtained for targets in nouns versus targets in verbs. When processing words that begin with stressed syllables, such as the nouns in this study, the segmentation that is initiated could facilitate detection of the first syllable by correctly isolating it on a first pass. This advantage could also facilitate the detection of the second syllable in such words by tentatively accessing the target-bearing word (and its cohorts), and thus providing a basis to anticipate the target phoneme in the second syllable. Conversely, the unstressed first syllables of the verbs would be at a disadvantage in terms of segmentation, as they would be treated as extensions of the preceding word, and thus would not be correctly isolated from the speech signal on a first pass. This would deprive their second syllables of the anticipatory processing advantage ascribed to nouns. Of course, it could be argued that the stressed second syllable of verbs have the same segmentation advantage as the first stressed syllable that is found in nouns. Such an effect might be expected to cancel out the category effect, and replace it with a stress effect. There are two alternatives to consider given the present data. Either Cutler and Norris' hypothesis must be abandoned, or we must suppose that the tentative lexical access advantage simply outweighs the segmentation advantage imparted by the strong second syllable in verbs. The resolution of this issue must be left to empirical examination.

The syntactic ambiguity effect can be approached in two ways. One way is to view the phenomenon as a decrement in the processing of ambiguous pre-target context. This approach is consistent with Foss's (1970) finding that subjects are slower in detecting phonemes in a monitoring task when they followed syntactically or lexically ambiguous contexts. Syntactic ambiguity was created in the current experiment primarily through the use of morphologically and syntactically (categorical) ambiguous words immediately preceding target-bearing words. This manipulation is quite similar to Foss's, which involved the use of lexically and referentially ambiguous words in pre-target positions.

The other approach is to treat the effect as a facilitation in the detection of syntactically unambiguous items. Buxton (1983) found that subjects show the serial position effects normally associated with the detection of stressed targets in a phoneme-monitoring task given stimuli in the form of either normal speech, or a form of jabberwocky in which the initial consonant of all content words in a normal sentence is replaced with another consonant to yield nonsense words. While this manipulation eliminated most of the semantic information in his stimulus sentences, it did not completely eliminate the syntactic information in the sentences. The preserved word-final and unbound morphemes in these sentences provide the listener with sufficient syntactic information to determine the syntactic categories of the target bearing words. In the present experiment, syntactic information seems to facilitate syllable detection. As stimuli with both ambiguous and non-ambiguous contexts are presented with normal and appropriate prosody, a purely metrical theory of syllable detection would not predict a difference between these two conditions. One interpretation of the result is that syntactic information provides a frame for anticipating the stress pattern of upcoming words in continuous speech. This frame could be used to guide stress-dependent aspects of the initial encoding of an anticipated word. This hypothesis is consistent with Grosjean and Gee's (1987) claim that words beginning with unstressed syllables may be accessed in the lexicon via their stressed syllables. Kelly and Bock (in press) analyzed a large corpus of bisyllabic, pure (non-category alternating) nouns and verbs, and found that 94% of these nouns took stress on the first syllable, while 69% of these verbs take stress on the second syllable. They also found that native speakers of English place stress on the first syllables of bisyllabic non-words in noun positions, and on the second syllables of non-words in verb positions in sentence frames. Given this distribution, a speech processor such as the one hypothesized by Grosjean and Gee could profit from a strategy of focusing processing resources on the stressed first syllable of

anticipated nouns, or the stressed second syllable of anticipated verbs. This line of thought is speculative, but we feel it invites further research.

Unfortunately, the current experiment does not address the distinction between the facilitation of processing unambiguous items, and the impedance of processing ambiguous items. It may be that the effect in this experiment is due to a combination of these factors. In any event, it appears that manipulations of syntactic ambiguity produce the same kind of effects that manipulations of target syllable stress do. This potentially restricts the usefulness of the syllable monitoring paradigm in examining the interaction of syntactic and metrical processes in the comprehension process.

In conclusion, it appears that data from the phoneme or syllable monitoring paradigm, while shedding light on the nature of factors effecting lexical access in speech recognition, cannot be interpreted simply in terms of metrical context effects. The role of syntactic ambiguity in monitoring provides reason to consider other methodologies for examining the role of stress in sentence processing which might not be sensitive to syntactic information.

EXPERIMENT 2

Given the interactions between phonological stress and other forms of linguistic context in determining reaction time on phoneme or syllable monitoring tasks, a second experiment was performed to determine if a post-sentence probe task would be sensitive to response latency effects that were more directly attributable to stress. There are several reasons to believe that a probe task might be an advance in this respect over monitoring tasks. The effect of syntactic or semantic ambiguity for instance, should be minimized in a probe task. As ambiguity is resolved during processing, its effects on the representation stored in short term memory should be eliminated. This would remove syntactic context effects which would otherwise mask phonological stress effects. The same argument holds for speech stream segmentation, or lexical access effects which depend on the listeners' tentative first pass attempts at processing speech input.

Another purpose behind this second experiment is to determine if phonological stress is represented in short term memory. Evidence from Taft and Hambly (1986), Grosjean (1985), and Cutler and Norris (1988) among others suggest that speech perception is not a simple left to right process like the one proposed by Marslen-Wilsen (1980). Context occurring both before and after a unit of speech may influence its processing. For this kind of contextual effect to be relevant in stress-related speech phenomena, it is essential that stress be represented in memory.

This issue is particularly significant in considerations of the psychological reality of the metrical theories of Liberman and Prince (1977), and Selkirk (1984). These theories argue that stress is assigned by a hierarchy defined over clauses, phrases or sentences. Such a hierarchy would require a listener to hold some sequence of speech in memory before stress assignments could be made within that sequence. Linguists might object to this view, claiming that the representation of stress, like other linguistic phenomena, is confounded by performance variables when it is examined in terms of speech production or speech perception. Even so, it would seem that the demonstration that stress is represented in short term memory is critical to assessing the validity of the data that linguists use to derive their theories. This concern is especially crucial, given Cooper and Eady's (1986) demonstration that speakers of a common dialect of English may have difficulty in reaching a consensus on how stress is assigned in a particular noun phrase.

Method

Materials. Experiment Two used the same auditory and visual stimuli that were used in Experiment One. In addition to these materials, Experiment Two introduced an additional word,"perfect" which shows the same alternation of stress position based on manipulations of syntactic category. Unlike the other words, perfect's stress alternation depends on the distinction between its use as an adjective and a verb. Like the other words, it appeared in ambiguous and unambiguous syntactic contexts, and appeared in forms with word-initial and word-final stress. These conditions are analogous to those employed with noun/verb homographs. In order to distribute the '5 experimental trials equally between blocks, a new trial order was constructed in which 5 experimental trials appeared in each of the last 5 blocks of the experiment. Unfortunately, the introduction of this stimulus caused a slight imbalance in the design. While each subject in Experiment One was given each probe position-sentence condition combination three times, subjects in this experiment were given one probe position-sentence condition four times. Despite this imbalance, an equal number of observations were made in each condition in the completed design.

Subjects. The subjects were 32 graduate and undergraduate students recruited from the Harvard University community. None of these subjects were in Experiment One. Of the 32, 15 were male, and 17 were female. All subjects were native speakers of English, with no discernable auditory or (uncorrected) visual deficits. Subjects were paid a base rate of $4.00 for their participation in the study, with an additional bonus of up to $2.00 paid on the basis of the speed and accuracy of their responses across all trials.

Procedure. Subjects were tested using the same set up as Experiment One. They were instructed to listen to sentences through a set of stereo headphones. The volume of the sentences was adjusted for each subject to ensure that they could hear all sentences clearly. As soon as each sentence ended, a tar[illegible] syllable appe[illegible]red on the screen in front of the subject, and remained there until the su[illegible] a response button on a mouse. Subjects were instructed to press the left button on the mouse if the sentence that they just heard included the syllable appearing on the screen, and the right button, if it did not. Reaction time and accuracy information were recorded for each response. Subjects received feedback, including their reaction time, the accuracy of their response, and the number of points they earned on the basis of their response towards a bonus fee following each of the first five trials. On the remaining trials, subjects only received accuracy feedback following incorrect responses. At the end of each twenty sentence block, subjects received summary feedback including their total number of errors, average reaction time, and the number of points they earned over the course of the block. Trials were arranged so that no experimental sentences appeared in the first block. Five experimental trials appeared randomly in each of the last five blocks. As was the case in experiment one, responses with latencies less than 500 msec., or greater than 2000 msec. were discarded.

Results

Subjects were able to perform the probe task with great accuracy. Probes were successfully detected on 96.08% of the experimental trials. Despite this high accuracy rate, one subject failed to detect all of the three stressed, second syllable targets that were probed for in contextually ambiguous verbs. To complete the design, this cell was filled in with the mean reaction time obtained in this condition (collapsing across stimulus) by the other 31 subjects.

The mean reaction times for all of the experimental conditions are presented in table 7. Both an ANOVA by subject, and an ANOVA by target-bearing word were performed. In contrast to the syllable monitoring

task, there was a significant main effect by subject for syllable stress in the probe task, $F(1,30) = 5.593$, $p = 0.023$. This effect was not significant when the data was analyzed by word, $F(1,24)$

= 2.898, p = 0.098. Stressed syllables were recognized faster than unstressed syllables for both word-initial and word-final targets. This was the only significant main effect or interaction in the analysis by subject. Analysis by word showed one significant effect. Word-initial targets were recognized faster than word-final targets, $F(1,24) = 5.261$, $p = 0.029$. When analyzed by subject, this comparison showed no significant effect, $F(1,31) = 0.101$, $p > 0.500$. Syntactic ambiguity had no significant effect on recognition latency in analysis by subject, $F(1,31) = 0.039$, $p > 0.500$, or analysis by word, $F(1,24) = 0.006$, $p > 0.500$.

Discussion

The results obtained in Experiment Two suggest that the probe task is sensitive to stress effects. Furthermore, the lack of effects for the syntactic ambiguity of target syllables at their time of presentation, and the position of target syllables in target-bearing words indicate that the probe task is not subject to the sources non-stress information which also contribute to facilitation in phoneme or syllable monitoring tasks.

This experiment was not designed to reveal the mechanism by which the probe task is sensitive to stress effects. Nevertheless, we would like to provide some speculation on the subject. An examination of the mean response times associated with correct responses to targets in different syllable positions in the stimulus sentences does not reveal evidence of a serial position effect that could be indicative of serial search. Given the lack of a serial search process, it is unlikely that the stress effect found on the probe task is attributable to the search mechanisms which seem to be responsible for the stress effects reported in the phoneme monitoring literature. Instead, we would like to suggest that these stress effects are attributable to the amount of processing that stressed syllables receive relative to unstressed syllables in speech recognition. Cutler and Norris' (1988) notion that strong syllables trigger segmentation, and are initially hypothesized to be the beginnings of lexical items in lexical access, suggests that stressed syllables are more actively processed or attended to than are unstressed syllables. Similarly, Grosjean and Gee's (1987) suggestion that stressed syllables are used to access words in the lexicon regardless of their position in a word suggests that stressed syllables receive deeper and earlier processing than do unstressed syllables. Once again, the available theories are suggestive, but there is still insufficient evidence to determine whether or not these factors are involved in the stress effect on the probe task.

General Discussion

Taken together, the results of Experiments One and Two suggest that the stress effects that are attributable to pretarget syntactic and lexical processing differ from those that are observable in a recognition task. This implies that stress, as it has been studied in the past, may be a conflation of these two separate effects. The perceptions of metrical phonologists observing their data, and psycholinguists designing their stimuli are based on recall (saying a word aloud, and then thinking about what it sounded like). Conversely, the effects that have been demonstrated in monitoring experiments rely on pretarget contextual processing. We are left then with the question of how much these effects overlap with one another. Do syllables in sentences receive the same prominence marking in anticipatory and recognition stress effects?

We have suggested that there are common mechanisms, such as segmentation by anticipated stressed syllables which may be play a role in the marking of both types of stress. Given this common mechanism, there is reason to expect at least some degree of overlap. The existence of such a mechanism though, is still an open empirical question. Nevertheless, the demonstration that different factors effect anticipatory and recognition stress effects suggests that this overlap may not be complete.

The probe task may be used to examine the relationship between these effects. Swinney's (1982) use of a lexical decision task to probe the time course of priming effects in sentence processing offers a model for how the probe task might be adapted for this end. The probe could be presented at different points in the sentence, to examine how the recognition stress effect develops over time as anticipatory processing is completed. One interesting use of this approach, would be to look for changes in recognition stress effects as a result of post-word context. Liberman and Prince (1977) argue that this context is important to the assignment of stress under conditions defined by the Stress Reduction Principle. Given the controversy over the data used to derive this principle (Cooper and Eady, 1986), this exploration could prove useful in examining the psychological reality of metrical phonology.

References

Ainsworth, W.A. (1972). Duration as a cue in the recognition of synthetic vowels. ***Journal of the Acoustical Society of America, 51***, 648-651.

Bernstein, L. E. (1983). Perceptual development for labeling words varying in voice onset time and fundamental frequency. ***Journal of Phonetics, 11***, 383-393.

Broadbent, D. E. (1958). ***Perception and Communication.*** London: Pergamon.

Buxton, H. (1983). Temporal predictability in the perception of English speech. In A. Cutler, & D.R. Ladd (Eds.), ***Prosody: Models and Measurements***, Berlin: Springer-Verlag, 111-122.

Chomsky, N. & Halle, M. (1968). ***The Sound Pattern of English***. New York: Harper and Row.

Cooper, W.E., & Eady, S.J. (1986). Metrical Phonology in Speech Production. ***Journal of Memory and Language, 25***, 369-384.

Cowan, N. (1984). On the short and long auditory stores. ***Psychological Bulletin, 96,*** 341-370.

Crowder, R. G. (1971). The sound of vowels and Consonants in immediate memory. ***Journal of Verbal Learning and Verbal Behavior, 10,*** 587-596.

Crowder, R. G. (1982). Decay of Auditory Memory in Vowel Discrimintation. ***Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,*** 153-162.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. ***Perception & Psychophysics, 20***, 55-60.

Cutler, A., & Foss, D. (1977). On the role of sentence stress in sentence processing. ***Language and Speech, 20***, 1-10.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. ***Journal of Experimental Psychology: Human Perception and Performance, 14 (1),*** 113-121.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. Cooper and E. Walker (Eds.), ***Sentence processing***. New York: Lawrence Erlbaum Associates.

Deutsch, J. A. & Deutsch, D. (1963). Attention: Some theoretical considerations. ***Psychological Review, 70***, 80-90.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. ***Psychological Review, 87,*** 272-300.

Foss, D. (1970). Some effects of ambiguity upon sentence comprehension. ***Journal of Verbal Learning and Verbal Behavior, 9***, 699-706.

Foss, D., & Lynch, R. (1969). Decision processes during sentence comprehension: Effects of surface structure on decision times. ***Perception and Psychophysics, 5***, 145-148.

Foss, D.J. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. ***Journal of Verbal Learning and Verbal Behavior, 8***, 547-562.

Fry, D.B. (1958). Experiments in the perception of stress. ***Language and Speech, 1***, 126-152.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. ***Language & Speech, 5,*** 171-189.

Fujisaki, H., & Kawashima, T. (1969). On the modes and mechanisms of speech perception. ***Annual Report of the Engineering Research Institute, Vol. 28,*** Faculty of Engineering, University of Tokyo, Tokyo, 67-73.

Fujisaki, H., & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. ***Annual Report of the Engineering Research Institute, Vol. 29***, Faculty of Engineering, University of Tokyo, Tokyo, 207-214.

Gilliom, J., & Sorkin, R. (1974). Sequential vs simultaneous two-channel signal detection: More evidence for a high-level interrupt theory. ***Journal of the Acoustical Society of America, 56,*** 157-164.

Glucksberg, S., & Cowen, G. N., Jr. (1970). Memory for nonattended auditory material. ***Cognitive Psychology, 1,*** 149-156.

Gordon, P. C., & Eberhardt, J. L. (1988, November). ***Effects of attention on the phonetic importance of cues***. Paper presented at the meetin of the Acoustical Society of America and Japan, Honolulu, Hawaii.

Gordon, P.C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. ***Perception & Psychophysics, 43***, 137-146.

Gordon, P.C. (in press). Context effects in recognizing syllable-final /z/ and /s/ in different phrasal positions. ***Journal of the Acoustical Society of America.***

Grodzinsky, Y. (1986). Language deficits and the theory of syntax. ***Brain and Language, 27(1),*** 178-191.

Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. ***Perception and Psychophysics, 38***, 299-310.

Grosjean, F., & Gee, J.P. (1987). Prosodic structure and spoken word recognition. ***Cognition, 25***, 135-155.

Hebb, D. (1949). ***The organization of behavior: A neuropsychological theory***. New York: Wiley.

Kelly, M.H., & Bock, J.K. (in press). Stress in time. Journal of ***Experimental Psychology: Human Perception and Performance.***

Klatt, D.H. (1976). Linguistic use of segmental duration in English: Acoustic and perceptual evidence. ***Journal of the Acoustical Society of America, 59***, 1208-1220.

Klatt, D.H. (1980a). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), ***Perception and production of fluent speech*** (pp.243-288).

Klatt, D.H. (1980b). Software for a cascade/parallel formant synthesizer. ***Journal of the Acoustical Society of America, 67***, 971-995.

Lehiste, I. (1970). ***Suprasegmentals***. Cambridge: MIT Press.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. ***Psychological Review, 74,*** 431-461.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. ***Journal of Experimental Psychology, 54,*** 358-368.

Liberman, A.M., Delattre, P.C., Gerstman, L.J., & Cooper, F.S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. ***Journal of Experimental Psychology, 52***, 127-137.

Liberman, M.Y. (1975). ***The intonational system of English***, Doctoral Dissertation, MIT, Cambridge, MA.

Liberman, M.Y., & Prince, A. (1977). On stress and linguistic rhythm. ***Linguistic Inquuiry, 8***, 249-336.

Lieberman, P. (1965). On the acoustic basis of perception of intonation by linguists. ***Word, 21 (1),*** 40-54.

Lieberman, P. (1967). Intonation, perception and language. ***Research Monograph no. 38***. Cambridge: MIT.

Lindholm, J. M., Dorman, M., Taylor, B. E., & Hannley, M. T. (1988). Stimulus factors influencing the identification of voiced stop consonants by normal- hearing and hearing-impaired adults. ***Journal of the Acoustical Society of America, 83,*** 1608-1614.

Marslen-Wilson, W.D. (1980). Speech understanding as a psychological process. In J.C. Simon (Ed.), ***Spoken language generation and understanding***. Dordrecht:Reidel,39-67.

Martin, J.G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. ***Psychological Review,79***, 487-509.

Massaro, D. W. (1970). Retroactive interference in short-term recognition memory for pitch. ***Journal of Experimental Psychology, 83***, 32- 39.

Mermelstein, P. (1978). On the relationship between vowel and consonant identification when cued by the same information. ***Perception & Psychophysics, 23***, 331-336.

Miller, G. A. (1956). The magic number seven, plus or minus two: Some limitations on our capacity for processing information. ***Psychological Review, 63***, 81-97.

Miller, J. (1981). Global precedence in attention and decision. ***Journal of Experimental Psychology: Human Perception and Performance, 9***, 1161-1174.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. ***Cognitive Psychology, 14***, 247-279.

Miller, J.L. (1987). Rate-dependent processing in speech perception. In A. Ellis (Ed.), ***Progress in the psychology of language (Vol. 3)***. Hillsdale, NJ: Erlbaum.

Miller, J.L. (1981). Effects of speaking rate on segmental distinctions. In P.D. Eimas & J.L. Miller (Eds.), ***Perspectives on the study of speech*** (pp. 39-74). Hillsdale, NJ: Erlbaum.

Miller, J.L., & Grosjean, F. (1981). How the components of speaking rate influence perception of phonetic segments. ***Journal of Experimental Psychology: Human Perception and Performance, 7***, 208-215.

Miller, J.L., & Liberman, A.M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. ***Perception & Psychophysics, 25***, 457-465.

Morton,J., and Long, J. (1976). Effect of word transitional probability on phoneme identification. ***Journal of Verbal Learning and Verbal Behavior, 15***, 43-51.

Nakatani, L.H., O'Connor, K.D., & Aston, C.H. (1981). Prosodic Aspects of American English Speech Rhythm. ***Phonetica, 38***, 84-106.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. ***Psychological Review, 86***, 214-255.

Neisser, U. (1967). ***Cognitive Psychology***. Appleton-Century-Crofts, New York.

Norman, D. A. (1968). Toward a theory of memory and attention. ***Psychological Review, 75***, 522-536.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. ***Psychological Review, 85***, 172-191.

Peterson, G.E., & Lehiste, I. (1960). Duration of syllabic nuclei in English. ***Journal of the Acoustical Society of America, 32***, Gordon, P.C. (in press). Context effects in recognizing syllable-final /z/ and /s/ in different phrasal positions. ***Journal of the Acoustical Society of America***. 693-703.

Pinker, S. (1984). ***Language learnability and language acquisition***. Cambridge: Harvard University Press.

Pisoni, D. B. (1971). ***On the nature of categorical perception of speech sounds***. Doctoral thesis, University of Michigan.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. ***Perception & Psychophysics, 13***, 253-260.

Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. ***Memory & Cognition, 3***, 7-18.

Pollack, I., & Pickett, J.M. (1963). The intelligibility of excerpts from conversational speech. ***Language and Speech, 6***, 165-171.

Port, R.F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. ***Journal of Phonetics, 7***, 45-56.

Port, R.F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. ***Perception & Psychophysics, 32 (2)***, 141-152.

Port, R.F., Al-Ani, S., & Maeda, S. (1980). Temporal Compensation and Universal Phonetics. ***Phonetica, 37***, 235-252.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), ***Information processing and cognition: The loyola symposium***. Hillsdale, NJ: Erlbaum.

Price, P. J., & Simon, H. J. (1984). Perception of temporal differences in speech by "normal-hearing" adults: Effects of age and intensity. ***Journal of the Acoustical Society of America, 76,*** 405-410.

Rakerd, B., & Fowler, C. (1984, October). The isochrony of speech does not appear to be syntactically conditioned. ***Paper presented at the meeting of the Michigan Linguistics Society***, East Lansing, MI.

Scharf, B. (1961). Complex sounds and critical bands. ***Psychological Bulletin, 58,*** 205-217.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. ***Psychological Review, 84,*** 1-66.

Selkirk, E.O. (1984). ***Phonology and syntax: The relationship between sound and structure***. Cambridge: MIT Press.

Shields, J.L., McHugh, A., & Martin, J.G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. ***Journal of Experimental Psychology, 102***, 250-255.

Shiffrin, R. M., & Dumais, S. T. (1981). The development of automatism. In J. R. Anderson (Ed.), ***Cognitive Skills and Their Acquisition.*** Hillsdale, NJ: Lawrence Erlbaum Associates.

Shiffrin, R. M., & Grantham, D. (1974). Can attention be allocated to sensory modalities? ***Perception & Psychophysics, 15,*** 460-474.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. ***Psychological Review, 84,*** 127-190.

Stevens, K. N., Liberman, A. M., & Ohlman, S. E. G. (1969). Cross-language study of vowel perception. ***Language & Speech, 12,*** 1-23.

Summerfield, Q. (1981). On articulatory rate and perceptual constancy in phonetic perception. ***Journal of Experimental Psychology: Human Perception & Performance, 7***, 1074-1095.

Sweet, H. (1908). ***The sounds of English- An introduction to phonetics***. Oxford: Clarendon Press.

Swinney, D. (1982). The effect and time-course of information interaction during speech comprehension:Lexical segmentation, access and interpretation. In J. Mehler, E. Walker, and M. Garrett (Eds.), ***Perspectives in mental representation***. Hilldale: Erlbaum.

Taft, M., & Hambly, G. (1986). Exploring the cohort model of spoken word recognition. ***Cognition, 22***, 259-282.

Tanner, W. P., Jr., Swets, J. A., & Green, D. M. (1956). Some general properties of the hearing mechanism ***(Technical Report, No. 30, Electron)***. Ann Arbor, Michigan: Defense Group, University of Michigan.

Treisman, A. M. (1964). Selective attention in man. ***British Medical Bulletin, 20,*** 12-16.

Treisman, A. M., & Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), ***Attention and performance IV***. New York: Academic Press.

Treisman, A. M., & Geffen, G. (1967). Selective attention: Perception or response? ***Quarterly Journal of Experimental Psychology, 19,*** 1-18.

Treisman, A. M., & Riley, J. (1969). Is selective attention selective perception or selective response? A further test. ***Journal of Experimental Psychology, 79,*** 27-34.

Verbrugge, R.R., & Shankweiler, D.P. (1977). Prosodic information for vowel identity. ***Journal of the Acoustical Society of America, 61***, S39 (Abstract).

Table 1

Percent Correct in Identifying Vowels

	Vowel			
Phrase Position	I	i	e	ai
Final	94.2	96.7	97.3	97.5
Internal	97.3	95.3	97.1	95.6

Table 2

Percent Correct in Identifying Fricatives

Fricative		Vowel			
		I	i	e	ai
	/s/	88.3	74.9	80.4	69.2
	/z/	96.6	96.5	97.9	96.5

Table 3

Formant Frequencies for Vowel Stimuli (300 ms & 50 ms)

Stimulus		Formant Frequencies (Hz)				
Number		F1	F2	F3	F4	F5
/i/	1	270	2300	3019	3500	4500
	2	285	2262	2960	3500	4500
	3	298	2226	2902	3500	4500
	4	315	2180	2836	3500	4500
	5	336	2144	2776	3500	4500
	6	353	2103	2719	3500	4500
/I/	7	374	2070	2666	3500	4500

Table 4

Mean Proportion /i/ Responses

	No-Distractor Condition			Distractor Condition		
Stimulus	Duration (Msec)					
Number	50	300	Mn.	50	300	Mn.
/i/ 1	1.000	1.000	1.000	.927	.964	.945
2	1.000	1.000	1.000	.881	.974	.927
3	.858	1.000	.929	.697	.948	.823
4	.317	.814	.565	.482	.844	.663
5	.033	.400	.217	.177	.661	.419
6	.017	.109	.063	.087	.330	.209
/I/ 7	.008	.025	.017	.092	.266	.174

Note: A vowel identification response is reported in the distractor condition only if an accurate response was made to the distractor task.

Table 5

STIMULUS TYPES USED IN EXPERIMENTS 1 AND 2

AMBIGUOUS CONTEXT

Noun Target

"Class **con**flicts give rise to revolution."

Verb Target

"Class con**flicts** with my two noon appointments."

UNAMBIGUOUS CONTEXT

Noun Target

"These **con**flicts cannot be avoided."

Verb Target

"Class often con**flicts** with lab meetings."

NOTE: Highlighting indicates stress. Both syllables in the target bearing word "conflict" act a targets in all four conditions.

TABLE 6

MEAN REACTION TIMES (MSEC.) FOR EXPERIMENT 1

UNAMBIGUOUS CONTEXT

Position	Stressed	Unstressed	Mn.
First	1378.14	1396.79	1387.46
Second	1291.60	1279.52	1285.56
Mn.	1334.87	1338.15	

AMBIGUOUS CONTEXT

Position	Stressed	Unstressed	Mn.
First	1419.17	1441.87	1430.52
Second	1366.00	1299.60	1332.80
Mn.	1392.58	1370.74	

TABLE 7

MEAN REACTION TIMES (MSEC.) FOR EXPERIMENT 2

UNAMBIGUOUS CONTEXT

Position	Stressed	Unstressed	Mn.
First	998.33	1023.43	1010.88
Second	942.56	985.17	963.87
Mn.	970.45	1004.30	

AMBIGUOUS CONTEXT

Position	Stressed	Unstressed	Mn.
First	977.07	992.85	984.96
Second	965.75	974.16	969.95
Mn.	971.41	983.50	

Figure Captions

Figure 1. Spectrograms for /i/ and /I/.

Figure 2. Proportion of /i/ responses as a function of duration and formant series.

Figure 3. Sequence of events on a trial. The top sequence shows a distractor trial and the bottom sequence shows a no-distractor trial.

Figure 4. Proportion of /i/ responses as a function of attention condition, duration, and formant series.

Figure 5. Proportion of /i/ responses for each formant series as a function of distractor task.

Figure 6. Proportion of /i/ responses for each duration as a function of distractor task.

Figure 7. Difference in proportion of /i/ responses between 50 msec stimuli and 300 msec stimuli as a function of attention condition.

Figure 8. Reaction time to the arithmetic task as a function of block number.

Figure 9. Proportion of /i/ responses as a function of attention condition, duration, and formant series. The top panel shows results from the first half of the session and the bottom panel shows results from the second half of the session.

Figure 1

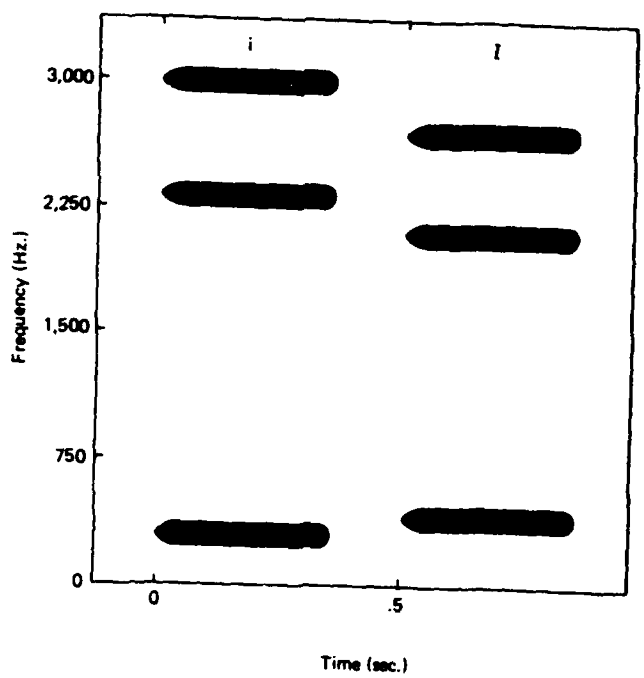
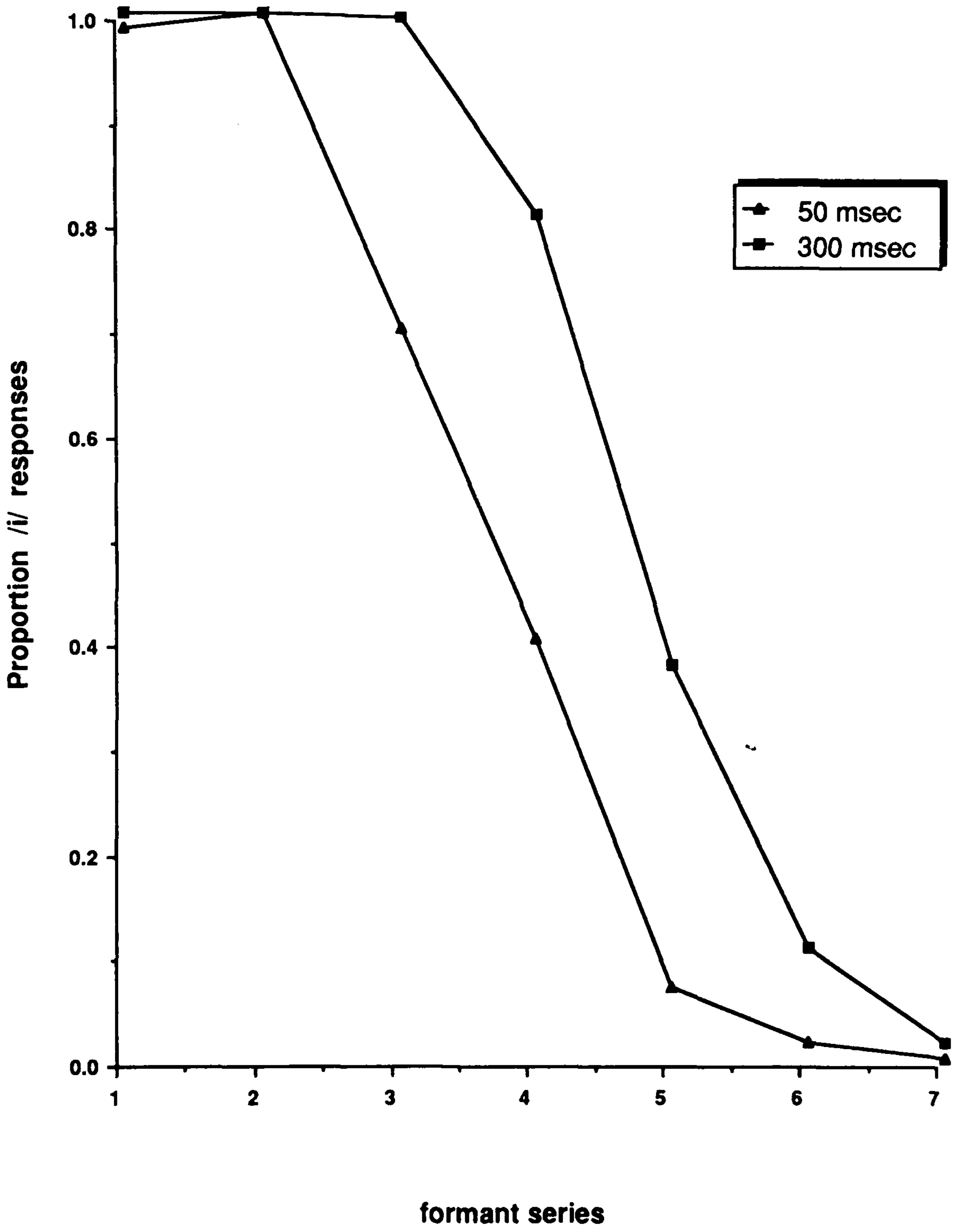
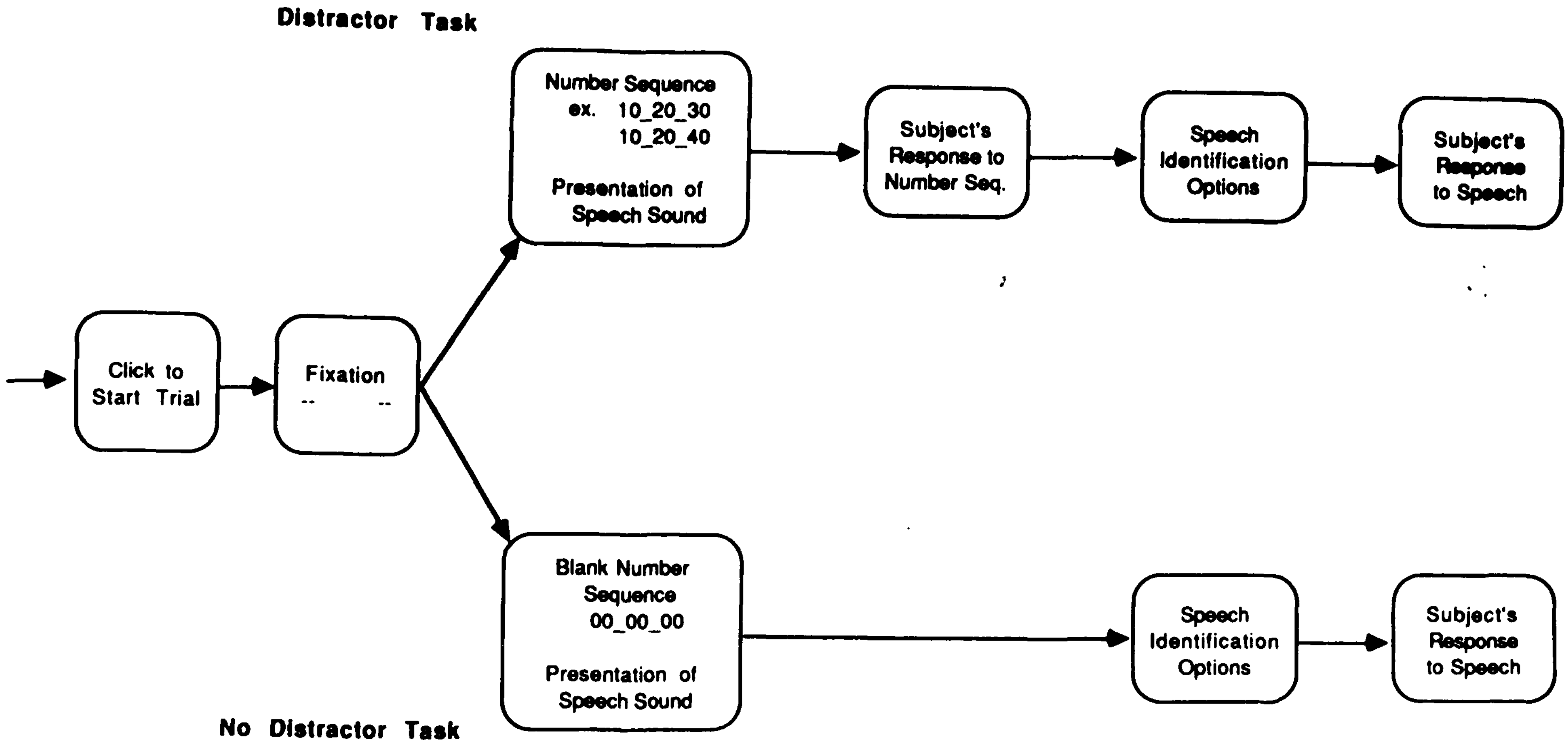


Figure 2



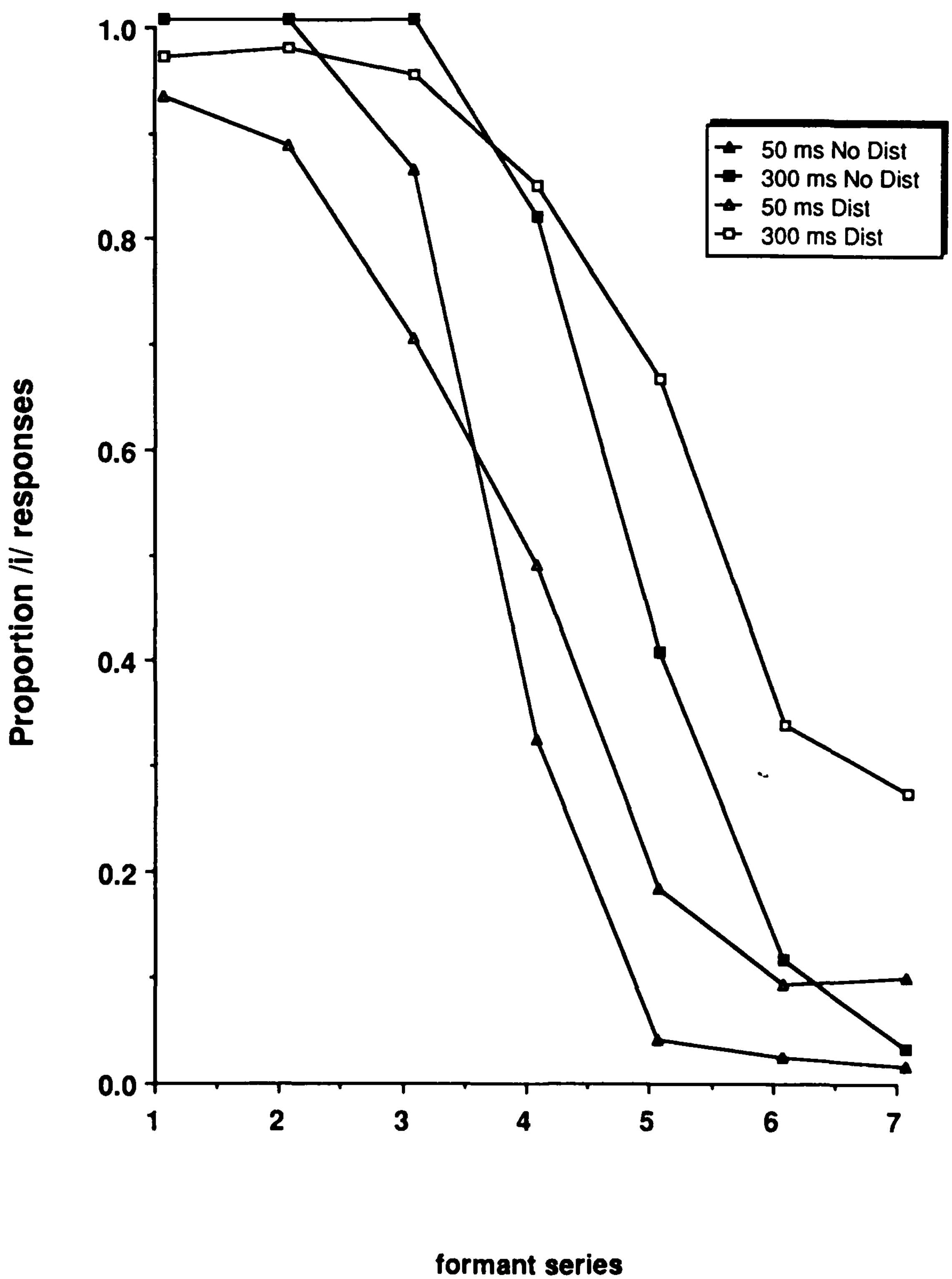
50 msec
300 msec
Proportion /i/ responses
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
formant series

Figure 3



Distractor Task
Click to Start Trial
Fixation
-- --
Number Sequence
ex. 10_20_30
10_20_40
Presentation of Speech Sound
Subject's Response to Number Seq.
Speech Identification Options
Subject's Response to Speech
Blank Number Sequence
00_00_00
Presentation of Speech Sound
Speech Identification Options
Subject's Response to Speech
No Distractor Task

Figure 4



50 ms No Dist
300 ms No Dist
50 ms Dist
300 ms Dist
Proportion /i/ responses
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
formant series

Figure 5

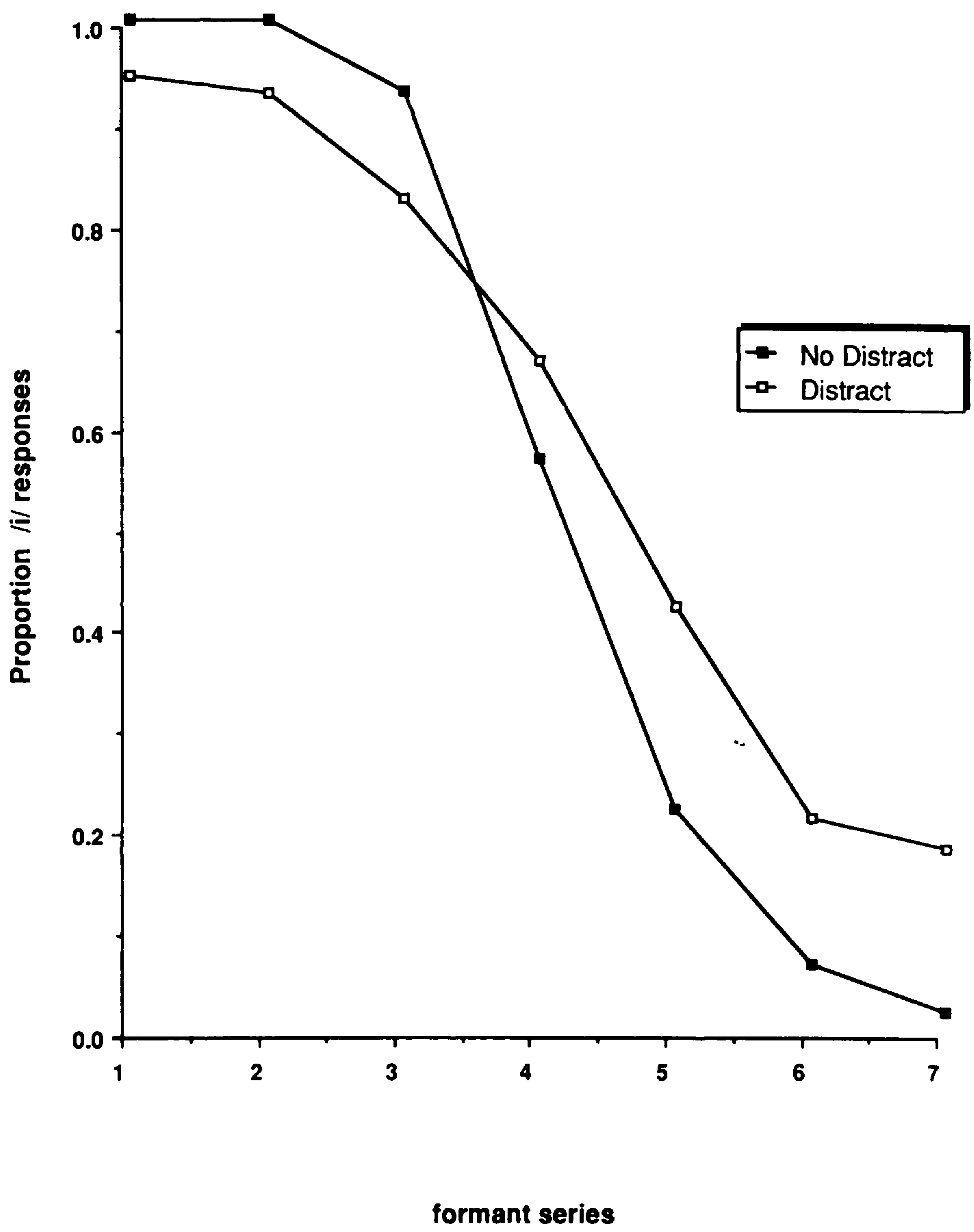

Proportion /i/ responses
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
No Distract
Distract
formant series

Figure 6

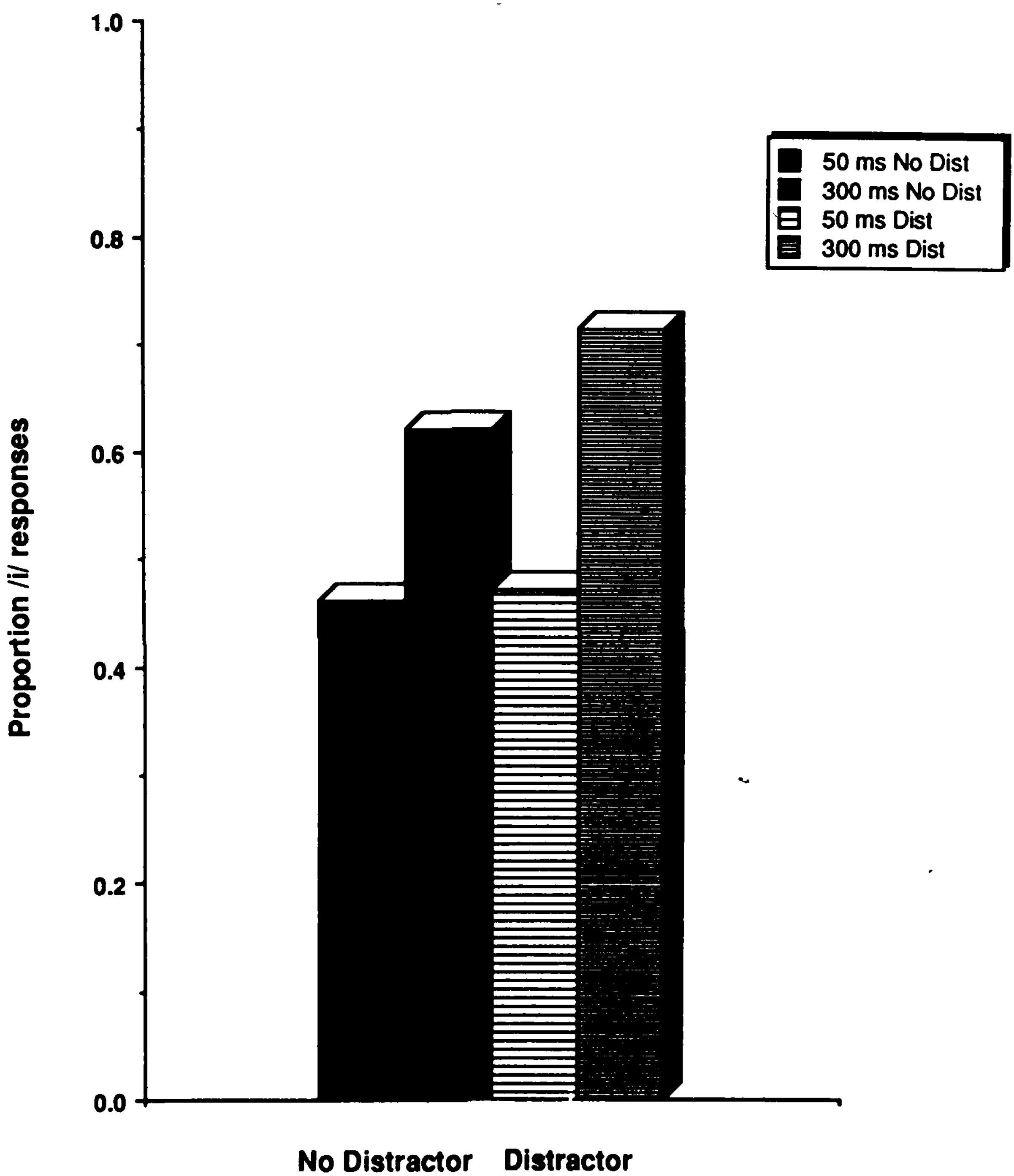
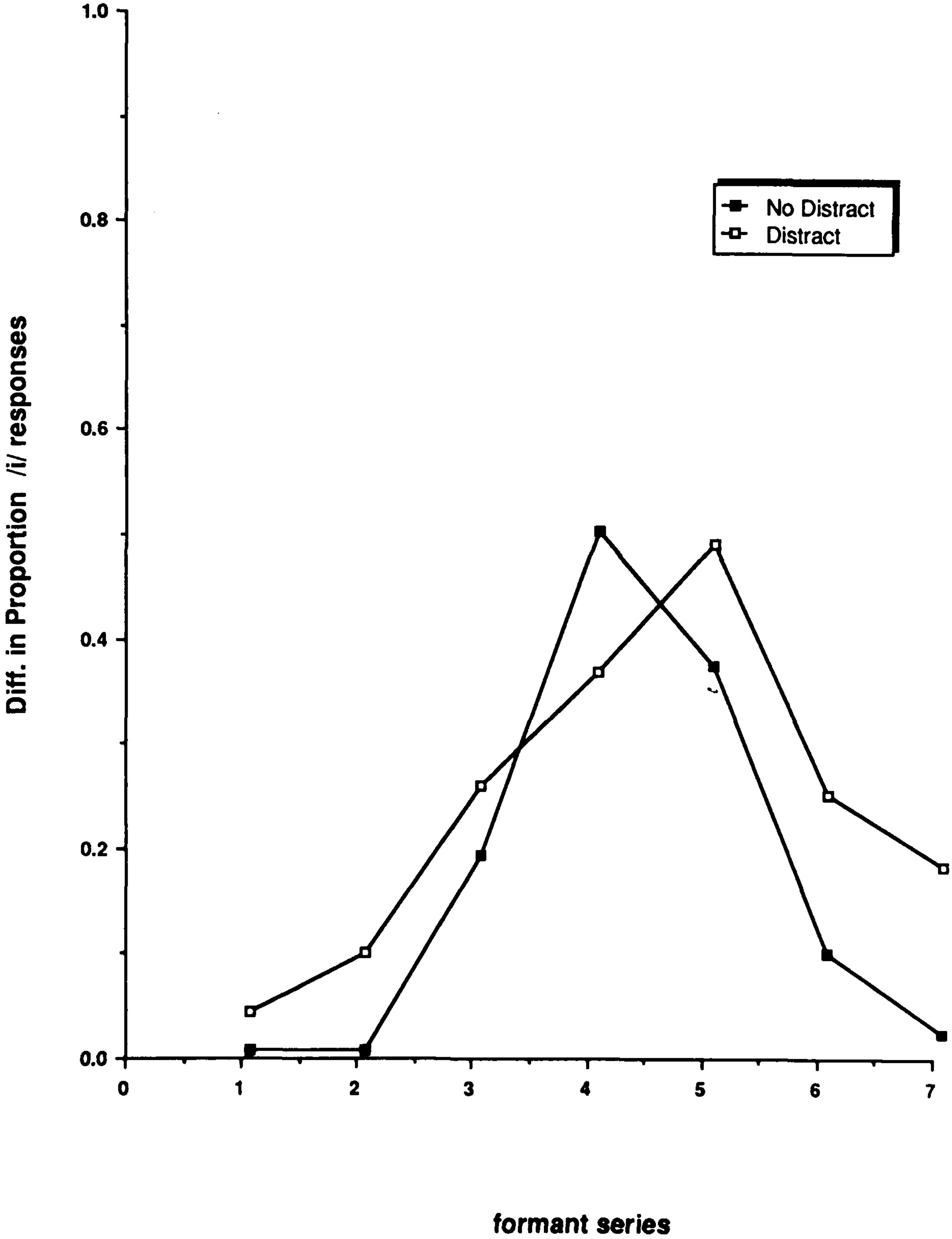

1.0
0.8
0.6
0.4
0.2
0.0
Proportion /i/ responses
50 ms No Dist
300 ms No Dist
50 ms Dist
300 ms Dist
No Distractor
Distractor

Figure 7



1.0
0.8
0.6
0.4
0.2
0.0
Diff. in Proportion /i/ responses
0
1
2
3
4
5
6
7
No Distract
Distract
formant series

Figure 8

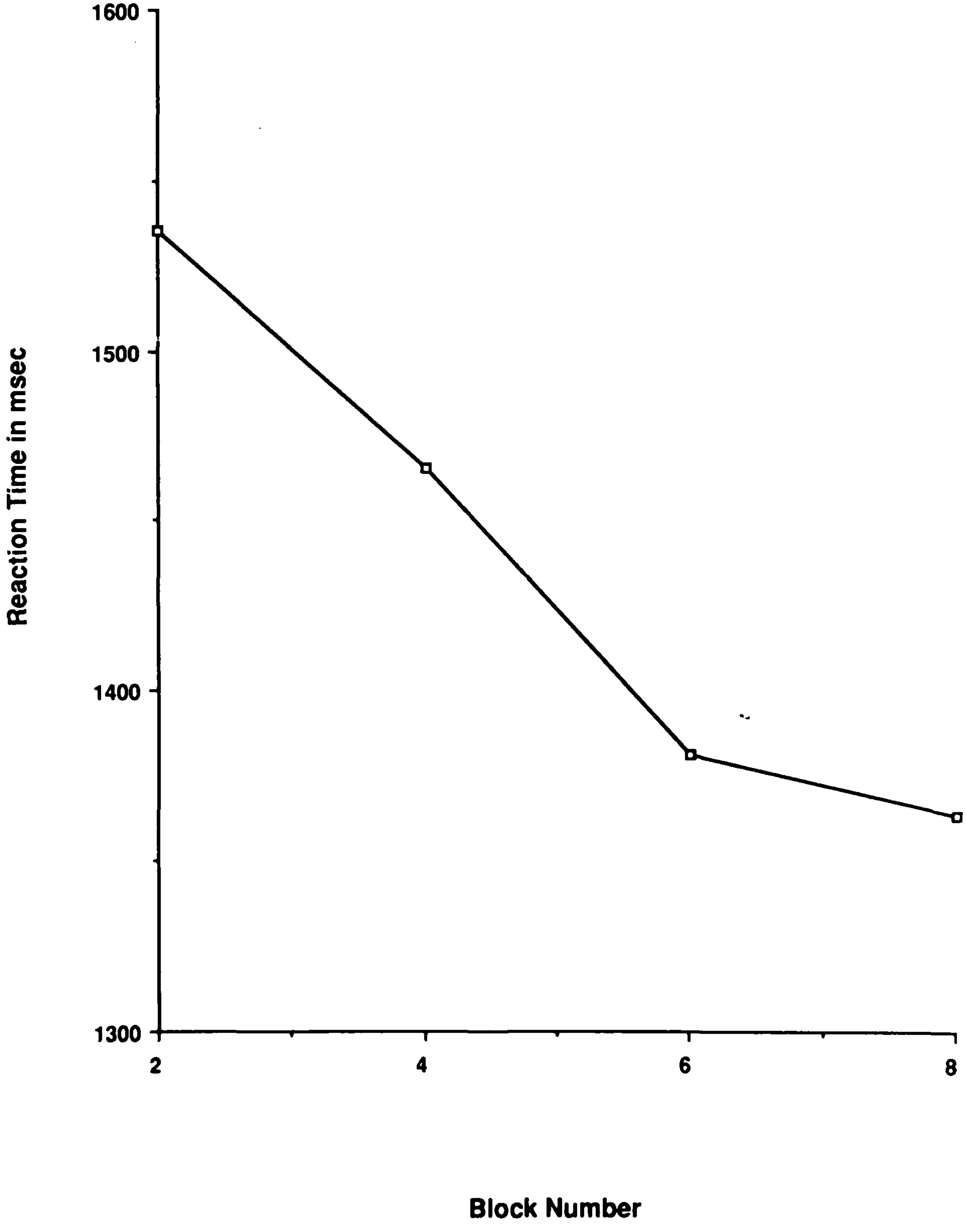
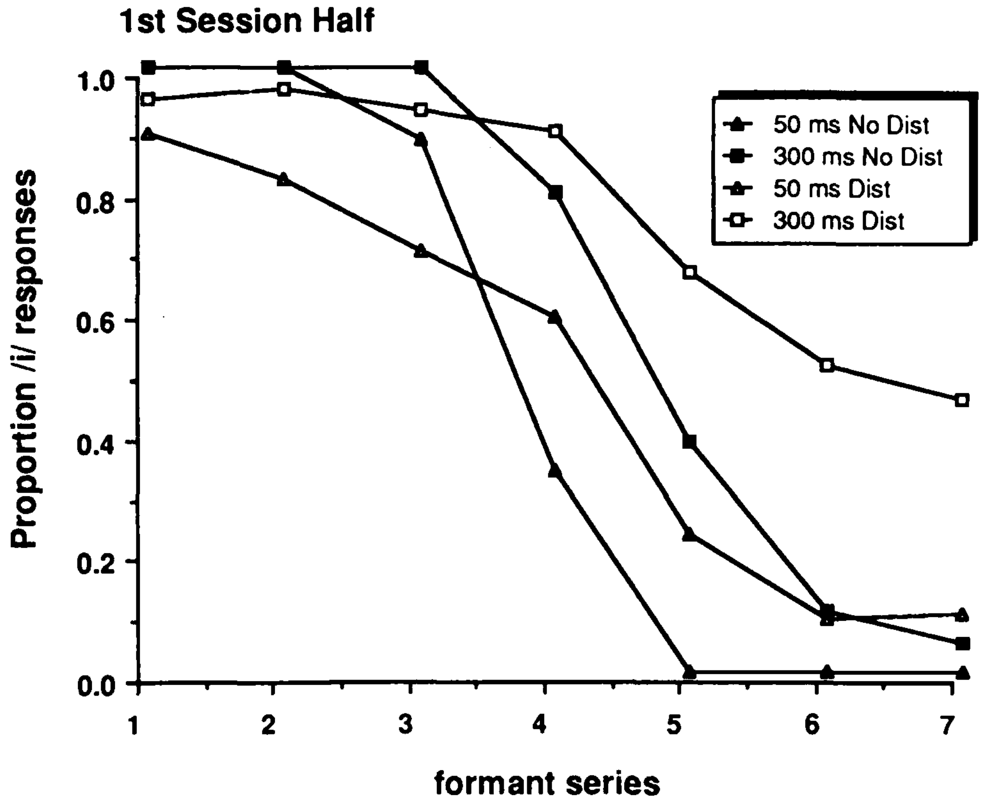
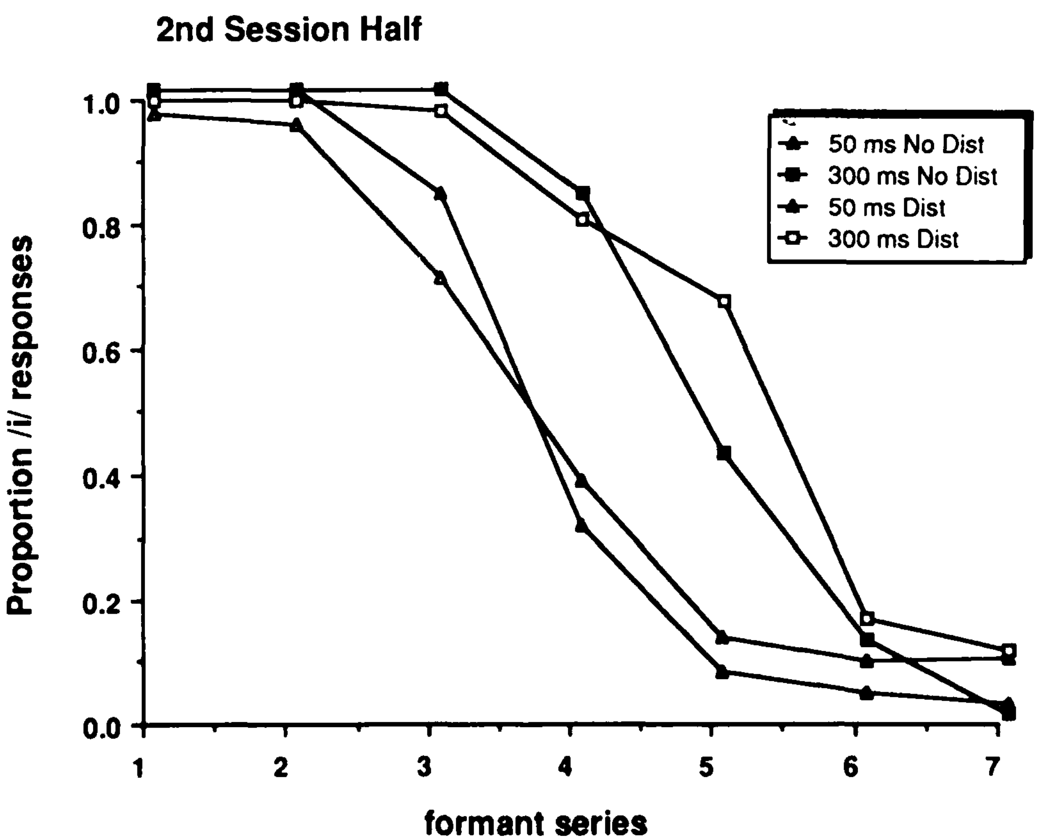


Figure 9


1st Session Half
Proportion /i/ responses
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
formant series
50 ms No Dist
300 ms No Dist
50 ms Dist
300 ms Dist


2nd Session Half
Proportion /i/ responses
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
formant series
50 ms No Dist
300 ms No Dist
50 ms Dist
300 ms Dist